



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

JULIANA MACHADO DA SILVA

**RELAÇÃO ENTRE ABUNDÂNCIA E LOCALIZAÇÃO DE  
DNAS REPETITIVOS E O TEMPO DE FLORESCIMENTO EM  
MILHO**

---

Londrina  
2025



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---



**IDR-Paraná**

Instituto de Desenvolvimento  
Rural do Paraná - IAPAR-EMATER



JULIANA MACHADO DA SILVA

**RELAÇÃO ENTRE ABUNDÂNCIA E LOCALIZAÇÃO DE  
DNAS REPETITIVOS E O TEMPO DE FLORESCIMENTO EM  
MILHO**

---

Londrina  
2025

JULIANA MACHADO DA SILVA

**RELAÇÃO ENTRE ABUNDÂNCIA E LOCALIZAÇÃO DE  
DNAS REPETITIVOS E O TEMPO DE FLORESCIMENTO EM  
MILHO**

Trabalho de Dissertação apresentado ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

Orientador: Prof. Dr. André Luis Laforga Vanzela.

Coorientador: Prof. Dr. Mateus Mondin.

Coorientador: Prof. Dr. Josué Maldonado Ferreira.

Londrina  
2025

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

da Silva, Juliana Machado.

RELAÇÃO ENTRE ABUNDÂNCIA E LOCALIZAÇÃO DE DNAS REPETITIVOS E O TEMPO DE FLORESCIMENTO EM MILHO / Juliana Machado da Silva. - Londrina, 2025.  
76 f.

Orientador: André Luís Laforga Vanzela.

Coorientador: Mateus Mondin.

Coorientador: Josué Maldonado Ferreira.

Dissertação (Mestrado em Genética e Biologia Molecular) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2025.

Inclui bibliografia.

1. DNA repetitivo - Tese. 2. Elementos transponíveis - Tese. 3. FISH - Tese. 4. Milhos compostos - Tese. I. Vanzela, André Luís Laforga . II. Mondin, Mateus . III. Maldonado Ferreira, Josué. IV. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular. V. Título.

CDU 575.1

JULIANA MACHADO DA SILVA

**RELAÇÃO ENTRE ABUNDÂNCIA E LOCALIZAÇÃO DE  
DNAS REPETITIVOS E O TEMPO DE FLORESCIMENTO EM  
MILHO**

Trabalho de Dissertação apresentado ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

**BANCA EXAMINADORA**

---

Orientador: Prof. Dr. André Luís Laforga  
Vanzela  
Universidade Estadual de Londrina - UEL

---

Profa. Dra. Lívia do Vale Martins  
Universidade Federal do Piauí - UFPI

---

Prof. Dr. Josué Maldonado Ferreira  
Universidade Estadual de Londrina - UEL

Londrina, 31 de janeiro de 2025.

## AGRADECIMENTOS

Aos meus pais, Vilza e Rozenildo, e aos meus irmãos, Eduardo e Guilherme, que me apoiaram em todas as decisões e, apesar da distância, compreenderam a importância dessa realização profissional. Sem vocês, eu jamais teria conseguido chegar aqui.

Ao meu orientador, Prof. Dr. André Vanzela, que contribuiu ativamente em minha formação e no início da minha jornada na Citogenômica Vegetal, área na qual me identifiquei e sou apaixonada desde então. Ao professor, Prof. Dr. Josué Ferreira, que possibilitou o estudo com os milhos compostos. Ao professor, Prof. Dr. Mateus Mondin pela concessão das sondas de K180 e TR-1 e conselhos de trabalho.

À Letícia Parteka, Rafael de Assis e Thaíssa Boldieri, que me acolheram desde a Iniciação Científica e, com paciência e boas risadas, me ensinaram tudo que sei hoje. À Thamirys Valentim, Rafaela Pinheiro, Karen Gomes, Guilherme Borges, Fernanda Treiber, Paola Chaves pela parceria de trabalho, risadas e amizade.

Ao meu namorado e parceiro de todas as horas, Ives Arenhart, que me apoia em todas as minhas decisões. Aos meus amigos, Giovana Gorini, Letícia Tartari e Willian de Paula, que são minha segunda família e os melhores amigos que eu poderia ter.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo auxílio financeiro. À Universidade Estadual de Londrina, que é a minha casa desde 2018 e possibilitou o meu crescimento pessoal e profissional.

"Y así encontraron la comida y ésta fue la que entró en la carne del hombre creado, del hombre formado; ésta fue su sangre, de ésta se hizo la sangre del hombre. Así entró el maíz en la formación del hombre por obra de los Progenitores. [...] De maíz amarillo y de maíz blanco se hizo su carne; de masa de maíz se hicieron los brazos y las piernas del hombre. Únicamente masa de maíz entró en la carne de nuestros padres"

*Popol Vuh. Las antiguas historias del Quiché*, traducidas do texto original com introdução e notas de Adrián Recinos, Colección Popular, núm. 11, FCE, México, 32ª reimp., 2005.

SILVA, Juliana Machado da. **Relação entre abundância e localização de DNAs repetitivos e o tempo de florescimento em milho**. 2025. 75 p. Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito para a obtenção do título de Mestre, Londrina, 2025.

## RESUMO

*Zea mays* é uma espécie modelo para diversos estudos citogenômicos, dada a proporção elevada de famílias de DNA que compõem sua fração repetitiva, sendo as principais o K180 (180pb) e o TR-1 (350 pb). Ambos podem formar blocos de heterocromatina e influenciar na recombinação genética e na expressão de genes próximos. O tempo de florescimento é um critério importante para o melhoramento vegetal, especialmente para o milho. Esse tempo parece estar, por sua vez, associado a variações no tamanho dos genomas e à presença de K180. Outra fração repetitiva relevante é a dos elementos de transposição, que compõem ~85% do genoma do milho. Esses elementos alteraram o arranjo das demais sequências nos genomas, e influenciam na expressão gênica e na evolução dos genomas. O objetivo do trabalho foi avaliar a diversidade cariotípica em cultivares de milhos compostos, utilizando o mapeamento de sequências repetitivas de DNA por FISH e bioinformática. As análises foram realizadas de maneira comparativa, com as linhagens NAM como referência. Para isso, mapeamos os cromossomos de diferentes cultivadores de milhos compostos utilizando DNAsats (K180, TR-1 e CentC) e DNAr 5S como sondas para FISH, além das estimativas dos valores 2C de DNA. Para as linhagens NAM, os pseudocromossomos foram mapeados por bioinformática em relação aos 14 principais genes da rota metabólica de florescimento. Os cariótipos dos milhos compostos exibiram heteromorfismos e polimorfismos para os sinais de FISH, embora apenas a sonda de DNAr 5S tenha mostrado sinais de FISH homogêneos. Por outro lado, os valores 2C de DNA não variaram de forma significativa. Com relação à análise dos pseudocromossomos, a distribuição de K180, TR-1 e CentC mostrou-se muito variável, tanto no número quanto no tamanho dos arranjos dessas duas sequências repetitivas. Neste caso, os arranjos de matrizes menores foram os mais diversos em posição e tamanho. A análise dos arranjos de DNAsats quanto à distância para os genes de florescimento, mostrou que há um número muito pequeno de monômeros e com as distâncias elevadas em relação aos genes alvo. Esses parâmetros não foram suficientes para explicar as diferenças no tempo de florescimento moduladas via efeito de posição dos arranjos de DNAsat. Quanto à abundância dos retrotransposons próximos aos genes *relacionados* ao tempo de florescimento, houve uma grande diversidade nas inserções gene-a-gene. De maneira geral, dois genes se destacaram pela quantidade relativa de inserções de RT-LTRs: *zfl2* e *mads1*. Dessa forma, o trabalho mostrou que a quantidade, tamanho e localização dos DNAsats (K180 e TR-1) em associação com o tamanho do genoma, não foi suficiente para estabelecermos uma correlação com o tempo de florescimento nos milhos compostos. Além disso, a anotação da coleção de sequências repetitivas próximas aos genes alvo mostrou que não há uma sustentação quanto aos efeitos de posição de K180 e TR-1 sobre o tempo de florescimento em milho. Contudo, a presença de fragmentos de TEs acerca desses genes e inseridos neles, sugere que estudos futuros de análises ômicas adicionais nos tecidos reprodutivos em milho sejam necessários, seja pela expressão dos genes alvo ou pela busca por RNAs não codificantes.

**Palavras-chave:** DNA repetitivo; Elementos transponíveis; FISH; Heteromorfismos; Milhos compostos.

SILVA, Juliana Machado da. **Relação entre abundância e localização de DNAs repetitivos e o tempo de florescimento em milho**. 2025. 75 p. Dissertação apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito para a obtenção do título de Mestre, Londrina, 2025.

## ABSTRACT

*Zea mays* is an ideal model for cytogenomic studies due to the high proportion of repetitive DNA families in its genome. The most prominent are the K180 (180 bp) and TR-1 (350 bp) families. Both can form heterochromatin blocks and influence genetic recombination and expression of nearby genes. Flowering time is a crucial criterion in plant breeding, especially in maize. This period seems to be associated with variations in genome size and the presence of K180. Another significant component of repetitive fraction is transposable elements (TEs), which constitute approximately 85% of the maize genome. These elements have reshaped the genomic landscape by altering the arrangement of other sequences, thereby influencing gene expression and genome evolution. This study aimed to evaluate the karyotypic diversity in composite maize varieties through the screening of repetitive DNA sequences by fluorescence *in situ* hybridization (FISH) and bioinformatics. The analyses were performed in a comparative framework, using the NAM lines as a reference point. For this purpose, we used DNAsats (K180, TR-1 and CentC) and 5S DNA probes for FISH, along the estimation of 2C DNA values. In the NAM lines, pseudochromosomes were mapped bioinformatically to 14 key genes in the flowering pathway. The composite maize karyotypes showed heteromorphisms and polymorphisms in FISH signals, although only the 5S rDNA probe produced homogeneous FISH signals. In contrast, the 2C DNA values remained relatively consistent. Pseudochromosome analysis revealed considerable variability in both the number and size of K180 and TR-1 arrays. The arrangements of smaller arrays showed the greatest diversity in position and size. An analysis of the DNAsat arrays in relation to their distance from flowering genes indicated a scarcity of monomers and a considerable distance from the target genes. These parameters were insufficient to explain the observed differences in flowering time based on position effects of the DNAsat arrays. Regarding the frequency of retrotransposons near genes associated with flowering time, we observed considerable diversity into gene-to-gene insertions. In general, two genes showed a remarkable prevalence of RT-LTR insertions: *zfl2* and *mads1*. Overall, our findings indicate that the number, size, and location of the DNAsats (K180 and TR-1) along with genome size were insufficient to establish a correlation with flowering time in composite maize. Furthermore, the annotation of the repetitive sequences in the vicinity of the target genes provided no evidence to support the hypothesis that the location of K180 and TR-1 influences flowering time in maize. Nevertheless, the presence of TE fragments near or within these genes suggests that future studies should include additional omics analysis in maize reproductive tissues, either by expression of the target genes or by identification of non-coding RNAs.

**Key-words:** FISH; Heteromorphisms; Maize composites; Repetitive DNA; Transposable elements.

## LISTA DE FIGURAS

- Figura 1** – Raças de milho ilustrando os tipos de grãos: (1). Pipocas da raça ‘Pisingallo’ colhidas na Argentina, (2). Milho doce da raça ‘Chullpi’ colhido no Peru, (3). Milho de sílex da raça ‘Sílex cubano’ colhido em Cuba, (4). Milho dentado da raça ‘Tuxpeño’ colhido no México, (5). Milho dentado da raça ‘Conico’ colhido no México, (6). Milho farinhento da raça ‘Cuzco’ coletado no Peru.....18
- Figura 2** – Variedades de milho NAM (*Nested Association Mapping*).....20
- Figura 3** – Classificação dos elementos de transposição .....24
- Figura 4** – Rota metabólica dos genes envolvidos no tempo de floração no meristema apical da folha e do broto .....28
- ARTIGO - Insights into the relationships between the distribution of repetitive DNA sequences and the flowering time of maize**
- Figure 1** – FISH on mitotic chromosomes with the K180 and CentC probes. The arrowheads indicate a secondary constriction. The asterisk indicates heteromorphism of chromosome pair 6. A) Mitotic metaphase of reference B73. B) Composite 1. C) Composite 2. D) Composite 4. E) Composite 6. F) Composite 7 .....61
- Figure 2** – FISH with TR-1, 5S DNAr and CentC probes. A) Metaphase of reference B73 with TR-1 and CentC probes. B) Metaphase of Composite 1 with CentC and 5S DNAr probes. C) Metaphase of Composite 1 with TR-1 and CentC probes. D) Metaphase of Composite 1 with TR-1 probes and arrows highlighting the small arrangements. E) Metaphase of Composite 2 with TR- 1 and CentC probes and arrowheads indicating the satellites of pair 6. F) Metaphase of Composite 4 with TR-1 and CentC probes. G) Metaphase of Composite 6 with the TR-1 probe. H) Metaphase of Composite 6 with TR- 1, CentC and 5S DNAr probes. I) Metaphase of Composite 7 with TR-1 and CentC probes. ....62
- Figure 3** – Idiograms with the localization sites of DNAsat K180 and CentC, comparing the reference B73 with the composite maize .....63
- Figure 4** – A) Number of K180 and TR-1 hybridization sites in composite and

NAM maize. B) 2C DNA value (pg) of the composite maize. Different letters indicate significant differences between treatments ( $p < 0.05$ ). Mean  $\pm$  S.E.;  $n = 3$ .....63

**Figuras suplementares**

**Supplementary Figure 1** – Idiograms of the pseudochromosomes of the NAM lines. The pair numbers are followed by the names of the genes related to the flowering time pathway they carry, and asterisks (\*) mark the position of each gene ..... 68

**Supplementary Figure 2** – Idiograms of the pseudochromosomes of the NAM lines (CML247, CML52, NC350, M37W, IL14H and P39). The pair numbers are followed by the names of the genes related to the flowering time pathway they carry, and asterisks (\*) mark the position of each gene. The pseudochromosomes have been centered according to the concentration of CentC arrays. Note that the late flowering lines accumulate more K180 than the others. However, in TR-1 the distribution of the blocks was more variable. Note also that the K180 on chromosome 9 (studied by Carvalho et al. 2022) is close to the *mads1* gene.....69

**Supplementary Figure 3** – Relative amount of K180 (A) and TR-1 (B) monomers in each pseudochromosome of NAM lineages .....70

**Supplementary Figure 4** – Abundance of Copia and Gypsy retrotransposons in the flanking regions (50,000 bp upstream and downstream) of the 14 genes related to flowering time of NAM B73. The chromosome number is shown on the left, followed by the gene analyzed. The grey shaded region corresponds to the gene region annotated from the information contained in MaizeGDB. The two blue bars in the analyzed gene regions mark the start and end of each gene .....71

**Supplementary Figure 5** – Abundance of Copia and Gypsy retrotransposons in the flanking regions (50,000 bp upstream and downstream) of the 14 genes related to flowering time of NAM maize lines, namely late CML247 (A) and CML52 (B), intermediate NC350 (C) and M37W (D) and early IL14H (E) and P39 (F). The chromosome number is shown on the left, followed by the gene analyzed. The grey shaded

region corresponds to the gene region annotated from the information contained in MaizeGDB. The two blue bars in the analyzed gene regions mark the start and end of each gene .....72

## LISTA DE TABELAS

<b>Supplementary Table 1</b> - Composite maize varieties and their respective reference. LMG: Laboratory for Genetic Improvement; C = Composites. ....	64
<b>Supplementary Table 2</b> - Characteristics of NAM strains and reference genomes. Accession codes at NCBI ( <a href="https://www.ncbi.nlm.nih.gov/">https://www.ncbi.nlm.nih.gov/</a> ) and information from Buckler et al. (2009). Phenotypic characteristics of strains ( <a href="https://www.maizegdb.org/genome/assembly/Zm-CML52-REFERENCE-NAM-1.0">https://www.maizegdb.org/genome/assembly/Zm-CML52-REFERENCE-NAM-1.0</a> ).....	65
<b>Supplementary Table 3</b> - Fourteen genes related to flowering, their respective positions in the maize chromosomes and the accession code in MaizeGDB. L = long arm and S = short arm. CL = Chromosome location.....	66
<b>Supplementary Table 4</b> - Distribution of FISH signals in composite landrace maize lineages. DNA 2C values (Mpb), number and size (L=large and S=small) of FISH signals (FS) for K180, TR-1 and CentC probes. MFT = Maize-Flowering Time.....	67
<b>Supplementary Table 5</b> - Distances less than 175 kbp from K180 to gene and number of K180 repeats at annotated position. Zfl1 = zea floriculata / leafy1. Dlf1: Delayed flowering 1. CL = Chromosome location. L = long arm. NR = Number of Repeats.....	67
<b>Supplementary Table 6</b> - Distances less than 175 kbp from TR-1 to gene and number of TR- 1 repeats at annotated position. Zfl1 = zea floriculata / leafy1. CL = Chromosome location. L= long arm. NR = Number of Repeats. ....	67

## LISTA DE ABREVIATURAS E SIGLAS

8 HQ	8-hidroxiquinolina
µg	Microgramas
µM	Micromolar
µL	Microlitro
BLASTn	<i>Basic Local Alignment Search Tool Nucleotide</i>
CENH3	<i>Centromere Specific Histone 3</i>
CentC	<i>Centromere repeat C</i>
CHROMO	<i>Chromatin Organization Modifier</i>
CRM	<i>Centromeric Retrotransposons of Maize</i>
CTAB	<i>Cetyltrimethylammonium bromide</i>
CV	Coeficiente de Variação
DAPI	4',6-diamidino-2-phenylindole
DNA	<i>Deoxyribonucleic acid</i>
DNAr	DNA ribossômico
DNAsat	DNA satélite
FISH	<i>Fluorescent in situ Hybridization</i>
IGS	<i>Intergenic Spacer</i>
kpb	Kilo pares de base
LINE	<i>Long Interspersed Nuclear Element</i>
LTR	<i>Long Terminal Repeats</i>
MITE	<i>Miniature inverted-repeat Transposable Element</i>
M	Molar
Mg	Miligrama

MgCl <sub>2</sub>	Cloreto de Magnésio
mL	Mililitro
Mmol/L	Milimole por litro
Mpb	Mega pares de base
NAM	<i>Nested Association Mapping</i>
NCBI	National Center for Biotechnology Information
ng	Nanogramas
PacBio	<i>Pacific Biosciences</i>
PCR	<i>Polymerase Chain Reaction</i>
pb	Pares de bases
PEG	polietilenoglicol
QTL	<i>Quantitative Trait Loci</i>
RCR	<i>Rolling-Circle Replication</i>
RNA	<i>Ribonucleic Acid</i>
RON	Região Organizadora de Nucléolo
RT-LTR	<i>Retrotransposons Long Terminal Repeat</i>
SSC	citrato de sódio salino
TE	Elemento de transposição
TSRT	<i>Target-Specific Reverse Transcription</i>
T2T	<i>telomere-to-telomere</i>
Vgt1	<i>Vegetative to Generative Transition 1</i>
V:v	Volume:volume
ZCN8	<i>Zea mays CENTRORADIALIS 8</i>
ZmCCT9	<i>CO CO-LIKE TIMING OF CAB 1 protein domain 2</i>

## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	14
2	<b>REVISÃO BIBLIOGRÁFICA</b> .....	16
2.1	<b>A importância do milho</b> .....	16
2.2	<b>A diversidade do milho</b> .....	16
2.3	<b>Mapeamento de sequências repetitivas em <i>Z. MAYS</i></b> .....	20
2.4	<b>Mapeamento cromossômico por FISH</b> .....	21
2.5	<b>Elementos de transposição no milho</b> .....	22
2.6	<b>Associação de dados genômicos e citogenéticos</b> .....	25
2.7	<b>O tempo de florescimento do milho e sua relação com DNA satélites</b> .....	26
2.8	<b>Conteúdo do genoma do milho</b> .....	28
3	<b>OBJETIVO GERAL E ESPECÍFICOS</b> .....	30
3.1	<b>Objetivo geral</b> .....	30
3.2	<b>Objetivos específicos</b> .....	30
4	<b>BIBLIOGRAFIA</b> .....	31
5	<b>ARTIGO</b> .....	37
6	<b>CONCLUSÃO</b> .....	63
7	<b>FIGURAS</b> .....	63
8	<b>MATERIAL SUPLEMENTAR</b> .....	67

## 1 INTRODUÇÃO

O milho (*Z. mays* L., Poaceae) foi domesticado a partir do (*Z. mays* ssp. *parviglumis*) e, posteriormente, sofreu introgressão de uma segunda espécie de teosinto (*Z. mays* ssp. *mexicana*). Após esses eventos, houve sua expansão pelas Américas. Como resultados de anos de seleção artificial, chegou-se à grande plasticidade fenotípica e adaptativa observada nas variedades cultivadas atualmente, desde as regiões temperadas até as tropicais. O Brasil é o terceiro maior produtor mundial de milho, com uma safra de mais de 115 mil toneladas colhidas em 2024, principalmente nas regiões tropicais e subtropicais do país. Com o avanço das mudanças climáticas, a produção do milho está instável, especialmente nas regiões próximas ao Equador, onde os impactos são mais severos. Portanto, o investimento na produção de variedades especiais e mais tolerantes às pressões bióticas e abióticas é crucial.

*Zea mays* é uma espécie modelo para diversos estudos citogenômicos, dada a proporção elevada de famílias de DNA que compõem sua fração repetitiva. Dentre essas, podemos citar: o Cent-C (*Centromere repeat C*), Cent-4, microsatélite TAG, B-específico, subtelômero, K180 (180 pb) e o TR-1 (350 pb). Há registros na literatura de que esses DNA satélites ocorrem desde grandes arranjos (blocos) até frações menores distribuídas pelos cromossomos. Ambos podem influenciar na recombinação genética e na expressão de genes próximos a essas regiões que são, em geral, heterocromáticas. As sondas de DNA para os *repeats* de K180 e TR-1, por exemplo, foram usadas para compreender a diversidade intra e interpopulacional de amostras de milho crioulo na Argentina. A citogenômica parece ser, portanto, uma excelente ferramenta de apoio ao melhoramento do milho, pois pode guiar a produção de novas variedades.

Os milhos de origem composta, que são o alvo desse estudo, carregam grande diversidade de marcas cromossômicas, incluindo polimorfismos e heteromorfismos. Cerca de 85% do genoma do milho é composto de sequências repetitivas, das quais a mais representativa é a dos elementos de transposição (TEs). Esses elementos podem agir de modo a alterar o arranjo das demais sequências nos genomas e influenciar na expressão gênica e na evolução dos genomas como um todo. O sequenciamento de moléculas longas permitiu reconhecer e anotar com maior qualidade e precisão as sequências completas dessas regiões repetitivas (TEs e DNAsats). Os cromossomos do milho foram sequenciados telômero-a-telômero, e a precisão foi de mais de 99,99% de cobertura do genoma total. Isso gerou uma riqueza de dados que permitiu explorar aspectos relacionados à proximidade entre diferentes famílias de DNA, à

genômica funcional, às associações entre genótipo e fenótipo, até a evolução desses genomas.

O tempo de florescimento, por exemplo, é um critério importante para o melhoramento vegetal, especialmente para o milho. Esse tem sido associado a variações no tamanho do genoma e à presença dos monômeros de K180. Variedades de milho com tempo de florescimento menor, em geral, têm tamanho de genoma menor e uma menor quantidade de K180. Essas informações abriram margem para avaliarmos a relação de todas as regiões com K180, incluindo grandes e pequenos arranjos próximos aos principais genes que fazem parte das rotas metabólicas do florescimento.

Neste estudo, utilizamos dois caminhos de análise: i) a diversidade cromossômica de cultivares de milhos compostos (não sequenciados), utilizando o mapeamento de sequências repetitivas de DNA por hibridização *in situ* e ii) o mapeamento comparativo de sequências repetitivas em genomas referência (linhagens NAM) completamente sequenciados de amostras não compostas. Esses dados foram analisados via bioinformática, relacionando-os ao tempo de florescimento estabelecido para cada variedade, onde organizamos as amostras em florescimento precoce, intermediário e tardio. Nesse sentido, foram identificados heteromorfismos cromossômicos baseados em sondas de K180, TR-1 e CentC, em milhos compostos, sempre comparados à referência B73, a qual é considerada de florescimento intermediário. Também relacionamos o tempo de florescimento dos milhos compostos com a abundância dos K180 e TR-1. Nossos dados gerados por análises *in situ* e *in situ* foram, por sua vez, comparados aos valores C de DNA, sob o enfoque do tempo de florescimento. Por fim, foram avaliadas a presença e a distância dos arranjos de K180, TR-1 e TEs inseridos ou adjacentes aos 10 genes mais importantes relacionados ao florescimento, nas linhagens de milho NAM.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 A IMPORTÂNCIA DO MILHO

O milho (*Z. mays* L., Poaceae) foi domesticado a partir do teosinto *Z. mays* ssp. *parviglumis* nas planícies do sudoeste do México, há cerca de 9.000 anos e, posteriormente, sofreu introgressão de uma segunda espécie de teosinto, *Z. mays* ssp. *mexicana*, em torno de 4.000 anos. Posteriormente, houve sua expansão nas Américas (KISTLER *et al.*, 2018). Atualmente, o milho está presente em todo o mundo, com uma grande variedade de grãos e sabores, o que o tornou um dos cereais mais cultivados e consumidos do mundo, tal como o trigo e o arroz (WANG *et al.*, 2018; GARCÍA-LARA e SERNA-SALDIVAR, 2019).

O milho tem grande relevância econômica, uma vez que é utilizado em diversos setores. Esse grão é consumido amplamente em receitas tradicionais, especialmente na América Latina e na África (REVILLA *et al.*, 2022). Além disso, parte da produção de milho é utilizada como ração no setor pecuário. O milho é matéria prima para produção de biocombustíveis, como alternativa à gasolina, sendo amplamente utilizado nos Estados Unidos (GUZZON *et al.*, 2021). Os Estados Unidos são o maior produtor mundial de milho, sendo responsável por 55% da sua produção, com uma produção anual que ultrapassa 340 milhões de toneladas (GARCÍA-LARA e SERNA-SALDIVAR, 2019; FAOSTAT, 2024). O Brasil é o terceiro maior produtor, com uma safra de mais de 115 mil toneladas colhidas em 2024 (USDA, 2024; CONAB, 2024). As mudanças climáticas tendem a impactar negativamente o rendimento do milho em países próximos ao Equador. Nessas regiões, o aumento contínuo das temperaturas favorecerá a maturação precoce da planta, produzindo menos biomassa nos grãos e, conseqüentemente, reduzindo a produtividade (FAROOQ *et al.*, 2023). Assim, investimentos na produção de variedades especiais que tolerem pressões bióticas e abióticas são cruciais.

### 2.2 A DIVERSIDADE DO MILHO

O milho possui uma grande diversidade de variedades, que foram obtidas por meio do melhoramento genético. Houve aumento de produtividade e adaptabilidade a uma diversidade de condições climáticas e de solo. Os tipos de milho dividem-se em crioulos ou nativos e comerciais. As variedades de milhos crioulos foram obtidas ao longo do tempo por pequenos e médios agricultores (EICHOLZ *et al.*, 2018). A vantagem dos milhos nativos está na sustentabilidade da produção como resistência a doenças, pragas e desequilíbrios climáticos, o

que mantém a diversidade genética das espécies, podendo servir de fonte para o melhoramento (CECCARELLI, 1994). O uso de sementes crioulas confere baixo custo de produção e permite a independência do produtor em comparação à necessidade de aquisição das sementes (SARAIVA *et al.*, 2019). Nesse sentido, os produtores familiares resgatam anualmente suas próprias sementes, obtendo plantas selecionadas nas condições edafoclimáticas impostas pelo nível socioeconômico do agricultor (SARAIVA *et al.*, 2019).

A América Latina possui uma rica diversidade de milhos crioulos. Esses são de grande valor patrimonial, ligados à produtos locais e tradicionais importantes para a soberania alimentar e identidades culturais (GUZZON *et al.*, 2021). As regiões de Chiapas, Mesa Central e Sierra Madre Ocidental são consideradas centros de diversidade de milho no México (PERALES e GOLICHER, 2014). Apesar disso, o maior registro de variedades de milho na América foi na Bolívia e no Peru, com 77 e 66 raças, respectivamente, superando o México (65 raças) (SERRATOS, 2009). Os tipos de milho mais comuns na América Latina são pipoca, doce, sílex, dentado e farinha (GUZZON *et al.*, 2021) (representados na Figura 1). No Brasil, o cultivo de milho crioulo é bastante comum, com destaque para as regiões Sudeste e Sul. O Estado de Santa Catarina, na região Sul do país, foi considerado um microcentro de diversidade de milho crioulo, onde foi descrito o, cultivo principalmente, de milho comum (dentados, semidentados, semiflintos e raramente Flint), doce e farináceo (COSTA *et al.*, 2016).

**Figura 1.** Raças de milho ilustrando os tipos de grãos: (1). Pipocas da raça ‘Pisingallo’ colhidas na Argentina, (2). Milho doce da raça ‘Chullpi’ colhido no Peru, (3). Milho de sílex da raça ‘Sílex cubano’ colhido em Cuba (4). Milho dentado da raça ‘Tuxpeño’ colhido no México, (5). Milho dentado da raça ‘Conico’ colhido no México, (6). Milho farinhento da raça 'Cuzco' coletado no Peru.



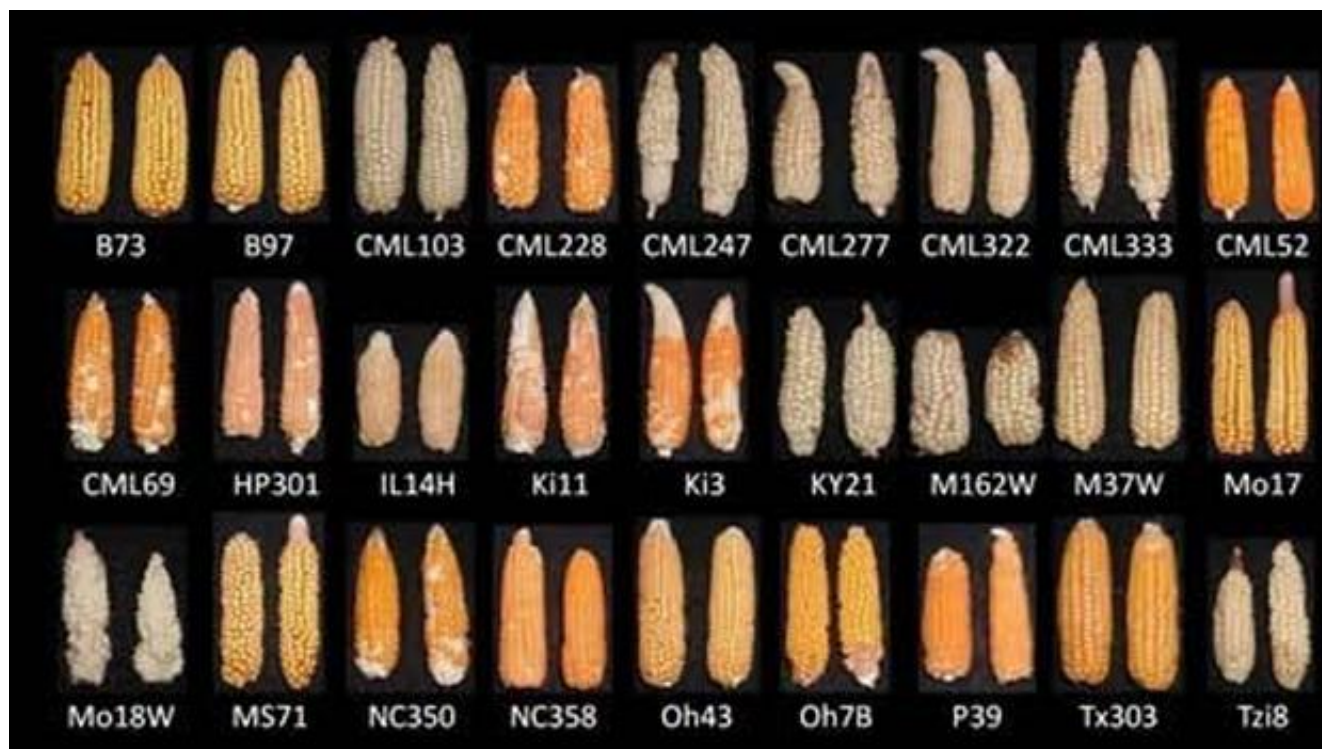
Fonte: Guzzon *et al.* 2022.

Os milhos comerciais, por sua vez, são cultivares híbridas patenteadas por empresas, para fins comerciais. O uso dessas tem como vantagens o aumento da produtividade, estabilidade e uniformidade das culturas (ANDRADE *et al.*, 2016). No entanto, a produção baseada nessas cultivares reduz o *pool* genético de germoplasmas comerciais, evidenciado pelo

fato de que as principais cultivares híbridas de milho no mundo vêm de cruzamentos envolvendo apenas sete linhagens: B73, Mo17, LH82, LH123, PH207, PH59 e PHG39 (MIKEL e DUDLEY, 2006). A variabilidade genética limitada dos atuais germoplasmas comerciais de milho pode levar ao aumento da vulnerabilidade à estresses abióticos e bióticos e, assim, resultar em insegurança alimentar e ganhos limitados com a seleção de culturas (ROMAY *et al.*, 2011).

Uma das estratégias para obter diversidade genética e novas variedades é via cruzamento de diversos parentais, a fim de produzir variedades compostas, concentrando, assim, características em uma só variedade (BOREM *et al.*, 2013). Esse é o caso dos milhos compostos, em que os melhoristas buscam cruzar variedades complementares a fim de aumentar a produtividade de grãos e conferir resistência a doenças foliares (KOLAWOLE *et al.*, 2018). Embora essa seja uma prática comum, são poucos os estudos de diversidade genética e citogenéticos focados na avaliação de compostos oriundos de parentais crioulos. Outro exemplo de cultivares aplicadas ao melhoramento são as NAM (*Nested Association Mapping*), formadas por um painel de cultivares criado a fim de auxiliar nos estudos de mapeamento genético de alta resolução de fenótipos complexos, tais como o tempo de floração e a resistência a doenças. Por meio desses estudos, foi possível elucidar, por exemplo, a regulação desses genes acima citados ([www.panzea.org](http://www.panzea.org); MCMULLEN *et al.*, 2009).

**Figura 2.** Variedades de milho NAM (*Nested Association Mapping*).



Fonte: McMullen *et al.* 2009.

### 2.3 MAPEAMENTO DE DNA SATÉLITES EM *Z. MAYS*

As sequências de DNA eucariótico estão presentes como cópias únicas no genoma ou como repetições, que podem ser intercaladas em todo o genoma ou repetidas em blocos (THAKUR *et al.*, 2021). As repetições intercaladas compreendem principalmente elementos transponíveis e seus fragmentos, que estão espalhadas por todo o genoma, enquanto as repetições em *tandem* são encontradas principalmente nas regiões proximais ou distais (GARRIDO-RAMOS, 2017; THAKUR *et al.*, 2021).

O milho é uma espécie amplamente estudada e utilizada como modelo para experimentos, dada a importância econômica. Em 1927, Barbara McClintock reportou pela primeira vez estruturas densamente coradas nos cromossomos, que foram popularizadas como *knobs*, e que atualmente sabemos que são ricas em monômeros de K180 (MCCLINTOCK, 1929; LONGLEY, 1941; DENNIS e PEACOCK, 1984). Dentre as sequências repetitivas comuns nos genomas de milho, podemos citar: o Cent-C (*Centromere repeat C*), Cent-4, microsatélite TAG, B-específico, subtelômero, K180 (180 pb) e o TR-1 (350 pb). O K180 é uma fração repetitiva em *tandem*, satélite, com monômeros de 180 pares de bases (pb). Essa fração aparece hipercondensada, com blocos variando de tamanho e posição nos cromossomos, podendo aparecer em heterozigose a depender da variedade analisada (DENNIS e PEACOCK,

1984; ANANIEV *et al.*, 1998). O TR-1 representa uma segunda família de DNA satélite, com monômeros com 350 pares de bases e, segundo ANANIEV *et al.* (1998b), pode ter derivado da repetição de 180 pb por duplicação. Uma região heterocromática hipercondensada pode ser composta apenas pela sequência de 180pb, por 350pb ou por ambos (ANANIEV *et al.*, 1998b). Estudos mais aprofundados a respeito do K180 demonstraram que esses monômeros podem influenciar na recombinação genética (STACK *et al.*, 2017) e na expressão de genes próximos a essas regiões que são, em geral, heterocromáticas (HABERER *et al.*, 2020). Normalmente, os K180 e TR-1 formam blocos nas regiões intersticiais a terminais dos cromossomos, de modo que esses DNAsats não comuns próximos aos centrômeros por meio da hibridização *in situ* (STACK *et al.*, 2017). Dados recentes de sequenciamento de moléculas longas sugerem que há arranjos menores, juntamente com esses monômeros em regiões proximais dos cromossomos de milho (dados não publicados).

Os centrômeros, por sua vez, são estruturas fundamentais para os eucariotos, pois durante a divisão celular, as cromátides-irmãs são movidas em direção aos polos das células por microtúbulos que se conectam às proteínas do cinetócoro, lateralmente aos centrômeros, regulando a dinâmica cromossômica. Os centrômeros funcionais são definidos epigeneticamente por histonas específicas de centrômero, as CENH3 (*Centromere Specific Histone 3*) (GENT *et al.*, 2018; HEUBERGER *et al.*, 2024). Nas plantas, os centrômeros possuem megabases de tamanho, compostos de sequências repetidas em *tandem* específicas. Em grande parte das gramíneas, as repetições em *tandem* se intercalam com RT-LTRs (*Retrotransposons Long Terminal Repeat*) pertencentes à superfamília Gypsy (HEUBERGER *et al.*, 2024). O milho possui centrômeros de um a dois mega pares de bases (Mpb) que são ricos em arranjos de CentC (*Centromere repeat C*) contendo 156 pb, além dos retrotransposons centroméricos (SCHNEIDER, *et al.*, 2016; ZHOU *et al.*, 2022). Tais sequências podem ser localizadas *in situ* e usadas para o mapeamento cromossômico (HUANG *et al.*, 2021; KIROV *et al.*, 2020).

#### 2.4 MAPEAMENTO CROMOSSÔMICO POR FISH

Uma técnica indispensável para o mapeamento cromossômico físico é a hibridização *in situ* fluorescente (FISH). Essa consiste na desnaturação e renaturação da fita de DNA, na presença de diferentes tipos de sondas marcadas ou diretamente conjugadas com fluorocromos (JIANG e GILL, 2006). Isso permite a localização de sequências de DNA específicas nos cromossomos e núcleos interfásicos. Dessa forma, a FISH pode ser utilizada como ferramenta de identificação cromossômica, como descrito em peônia (*Paeonia* L., Paeoniaceae) (CUI *et*

*al.*, 2022) e alface (*Lactuca sativa* L., Compositae) (WIDARMI *et al.*, 2020), dos níveis de ploidia, como reportado em amendoim (*Arachis hypogaea* L., Fabaceae) (ZHANG *et al.*, 2016), bem como em análises filogenéticas, como descrito em lírio (*Lilium ssp.*, Liliaceae) (ZHOU *et al.*, 2020).

Apesar da estabilidade no número cromossômico do milho ( $2n = 20$ ), exceto pela ocorrência de cromossomos B, regiões cromossômicas específicas em seus cariótipos são variáveis. KATO *et al.* (2004), por exemplo, identificaram, individualmente, os 10 pares cromossômicos em linhagens de milho, mediante pintura cromossômica usando diferentes sequências repetitivas como sondas para FISH, como DNA ribossomal, centromérico, telomérico e repeats. Os autores descreveram polimorfismos nos pares cromossômicos 2-4 e 7-9 dentro das linhagens analisadas em questão. O emprego de diversas sondas de DNAsat na FISH é uma maneira eficaz para a identificação de cada par cromossômico das variedades e subespécies de milho. Contudo, a pintura cromossômica baseada em sequências de DNA repetitivo é limitada a esse grupo vegetal (KATO *et al.* 2004; ALBERT *et al.*, 2018). Técnicas citogenômicas mais modernas utilizando oligonucleotídeos sintetizados como sondas para FISH (oligo-FISH) permitiram, por outro lado, uma maior versatilidade na identificação cromossômica em *Zea*, gerando um código de barras individual para cada cromossomo do gênero (JIANG, 2019; BRAZ *et al.*, 2020).

As sondas de DNA mais utilizadas para a identificação dos pares cromossômicos em milho são K180, TR-1, CentC, Cent4 e o DNA subtelomérico. As sondas K180 e TR-1 foram, recentemente, usadas para compreender a diversidade intra e interpopulacional de amostras crioulas de milho argentino (REALINI *et al.*, 2018). Variações nas frações de DNAsat foram correlacionadas de forma proporcional com altitude de cultivo; variedades que tipicamente ocorrem em regiões elevadas, tendem a correlacionar-se positivamente com a sonda de K180 e negativamente com TR-1 (FOURASTIÉ *et al.*, 2018). Estudos de localização de regiões centroméricas com a sonda CentC, mostraram diferenças claras no tamanho dos sinais de hibridização. KATO *et al.* (2004) sugeriram que esse polimorfismo deve-se a diferenças no número de cópias desses monômeros entre variedades, podendo igualmente serem úteis na identificação dos seus cromossomos. Dada a importância do desenvolvimento de mapas físicos cromossômicos, unir essa ferramenta de análise com o melhoramento de culturas poderia ser útil para guiar a produção de novas variedades. O uso dessa técnica é importante na busca por parentais em situações de cruzamento, especialmente quando se trata de cultivares compostas de milho consideradas antigas ou com *pool* de parentais “indeterminados”. Os milhos

compostos, em geral, carregam grande diversidade de marcas cromossômicas, polimorfismos e heteromorfismos (dados não publicados).

## 2.5 ELEMENTOS DE TRANSDUÇÃO NO MILHO

Os TEs, descobertos em mutantes de milho com grãos variegados (FESCHOTTE, 2023), podem agir de modo a alterar a expressão gênica e influenciar a evolução de todo o genoma. De acordo com HASSAN *et al.* (2024), inserções de TEs em diferentes regiões dos genomas, interrompendo ou restaurando de funções *upstream* e *downstream* de regiões gênicas, por exemplo, podem ter efeitos consideráveis na estabilidade dos genomas. Dentro da grande diversidade de TEs que compõem um genoma, há alguns tão estáveis que podem permanecer em locais cruciais para o funcionamento da cromatina, como os CRM (*Centromeric Retrotransposons of Maize*) descritos para milho (JIN *et al.*, 2004).

A filogenia dos TEs é complexa, mas, apesar disso, foi possível classificá-los de acordo com os seus mecanismos de movimento no genoma. Eles foram divididos nas Classes I e II (Figura 1). Os elementos de Classe I (retrotransposons) são transcritos em RNA, que por sua vez é revertido em DNA e inserido no genoma, em sequência (*cis*) ou em outro local (*trans*) em relação ao elemento de origem. Já os elementos de Classe II (transposons de DNA) utilizam outros mecanismos para sua inserção em outro local do genoma. A transposase, típica dos transposons, faz com que o elemento seja excisado na origem e reinserido em um novo local (WICKER *et al.*, 2007). Dentro dos elementos de Classe I, há os RT-LTRs, que são divididos nas seguintes superfamílias: Copia, Gypsy, Bel-Pao, Retrovirus e ERV (WICKER *et al.*, 2007). Os RT-LTR podem ser autônomos e ativos ou não autônomos, devido a degenerações, ou seja, os autônomos possuem a estrutura completa (com todos os domínios codificadores de proteínas para a mobilização) e os não autônomos estão incompletos (HASSAN *et al.*, 2024). Os elementos sem RT-LTRs ou LINEs (*Long Interspersed Nuclear Element*) também se transpõem via RNA intermediário. Porém, utilizam um mecanismo conhecido como *Target-Specific Reverse Transcription* (EICKBUSH e MALIK, 2007).

**Figura 3.** Classificação dos elementos de transposição.

Classification		Structure	TSD	Code	Occurrence
Order	Superfamily				
<b>Class I (retrotransposons)</b>					
LTR	<i>Copia</i>	→ GAG AP INT RT RH →	4-6	RLC	P, M, F, O
	<i>Gypsy</i>	→ GAG AP RT RH INT →	4-6	RLG	P, M, F, O
	<i>Bel-Pao</i>	→ GAG AP RT RH INT →	4-6	RLB	M
	<i>Retrovirus</i>	→ GAG AP RT RH INT ENV →	4-6	RLR	M
	<i>ERV</i>	→ GAG AP RT RH INT ENV →	4-6	RLE	M
DIRS	<i>DIRS</i>	→ GAG AP RT RH YR ←	0	RYD	P, M, F, O
	<i>Ngaro</i>	→ GAG AP RT RH YR → → →	0	RYN	M, F
	<i>VIPER</i>	→ GAG AP RT RH YR → → →	0	RYV	O
PLE	<i>Penelope</i>	← RT EN →	Variable	RPP	P, M, F, O
LINE	<i>R2</i>	RT EN	Variable	RIR	M
	<i>RTE</i>	APE RT	Variable	RIT	M
	<i>Jockey</i>	ORFI APE RT	Variable	RIJ	M
	<i>L1</i>	ORFI APE RT	Variable	RIL	P, M, F, O
	<i>I</i>	ORFI APE RT RH	Variable	RII	P, M, F
SINE	<i>tRNA</i>		Variable	RST	P, M, F
	<i>7SL</i>		Variable	RSL	P, M, F
	<i>5S</i>		Variable	RSS	M, O
<b>Class II (DNA transposons) - Subclass 1</b>					
TIR	<i>Tc1-Mariner</i>	→ Tase* ←	TA	DTT	P, M, F, O
	<i>hAT</i>	→ Tase* ←	8	DTA	P, M, F, O
	<i>Mutator</i>	→ Tase* ←	9-11	DTM	P, M, F, O
	<i>Merlin</i>	→ Tase* ←	8-9	DTE	M, O
	<i>Transib</i>	→ Tase* ←	5	DTR	M, F
	<i>P</i>	→ Tase ←	8	DTP	P, M
	<i>PiggyBac</i>	→ Tase ←	TTAA	DTB	M, O
	<i>PIF-Harbinger</i>	→ Tase* ORF2 ←	3	DTH	P, M, F, O
	<i>CACTA</i>	→ Tase ORF2 ←	2-3	DTC	P, M, F
	Crypton	<i>Crypton</i>	→ YR ←	0	DYC
<b>Class II (DNA transposons) - Subclass 2</b>					
Helitron	<i>Helitron</i>	→ RPA // Y2 HEL ←	0	DHH	P, M, F
Maverick	<i>Maverick</i>	→ C-INT ATP // CYP POL B ←	6	DMM	M, F, O

Structural features					
→	Long terminal repeats	←	Terminal inverted repeats	█	Coding region
—	Diagnostic feature in non-coding region	—	Region that can contain one or more additional ORFs	—	Non-coding region
Protein coding domains					
AP, Aspartic proteinase	APE, Apurinic endonuclease	ATP, Packaging ATPase	C-INT, C-integrase	CYP, Cysteine protease	EN, Endonuclease
ENV, Envelope protein	GAG, Capsid protein	HEL, Helicase	INT, Integrase	ORF, Open reading frame of unknown function	
POL B, DNA polymerase B	RH, RNase H	RPA, Replication protein A (found only in plants)		RT, Reverse transcriptase	
Tase, Transposase (* with DDE motif)		YR, Tyrosine recombinase		Y2, YR with YY motif	
Species groups					
P, Plants	M, Metazoans	F, Fungi	O, Others		

Fonte: WICKER *et al.* (2007).

Cerca de 85% do genoma do milho é composto de TEs, sendo os retrotransposons correspondentes à maior fração (SCHNABLE *et al.*, 2009; STITZER *et al.*, 2021), com mais de 140 mil sequências correspondentes aos RT-LTRs (Copia, Gypsy e *Unknown* LTR). Cento e setenta mil sequências correspondem aos transposons, sendo 20 mil sequências foram identificadas como Helitrons (STITZER *et al.*, 2021). Os retrotransposons que predomina na região centromérica de milho, chamados de CRM, fazem parte de um clado conhecido como cromovírus da superfamília Gypsy (ZHONG *et al.*, 2002). Esses elementos são constituídos de um cromodomínio (domínio *Chromatin Organization Modifier* - CHROMO) ou um domínio de direcionamento, que possui afinidade com a CENH3, conhecido como *CR motif* (*Centromeric Retrotransposons*), seguidos da integrase (HOUBEN *et al.*, 2007; NEUMANN *et al.*, 2011). Contudo, a maior parte dos RT-LTRs aparece distribuído ao longo dos cromossomos, às vezes formando pequenos arranjos (ASSIS *et al.*, 2020). Os avanços recentes acerca da caracterização e filogenia da fração repetitiva de DNA aconteceram não apenas em função dos sequenciamentos de alta qualidade, novas ferramentas de bioinformática e pela compilação de bancos de dados mais curados, mas também pelo aprimoramento de uso de sondas para FISH.

## 2.6 ASSOCIAÇÃO DE DADOS GENÔMICOS E CITOGENÉTICOS

O excesso de dados produzidos na área da Biologia Molecular e genômica conduziu à criação de um sistema computadorizado de coleta, organização e análise global de informações (GOMEZ-CASATI *et al.*, 2018). Há inúmeras ferramentas de bioinformática que fornecem sistemas rápidos e eficazes de gerenciamento de dados acerca de sequências de nucleotídeos e aminoácidos, como também de domínios e estruturas de proteínas e padrões de expressão para diversos organismos (TAN *et al.*, 2022). O sequenciamento de moléculas longas trouxe qualidade e quantidade de informações, de modo que as anotações de genes e regiões cromossômicas estão muito precisas, especialmente para plantas modelo. A plataforma MaizeGDB (WOODHOUSE *et al.*, 2021), por exemplo, possui diversos dados de referência, que são muito úteis para comparações em acessos ainda alvo de validação (GOMEZ-CASATI *et al.*, 2018).

Essas ferramentas permitem a produção de sondas de alta precisão de frações repetitivas de DNA, como TEs e DNAsats (SAHU e LIU, 2023). No feijão-azuki, *Vigna angularis*, LI *et al.* (2024), assim como no tomate selvagem, *Solanum pimpinellifolium* (HAN

*et al.*, 2024), por exemplo, há dados precisos da distribuição de sequências repetitivas para cada um dos pseudocromossomos das espécies em questão. O sequenciamento dos genomas de milho também foi amplamente explorado, dada a sua importância econômica. A variedade de milho B73 foi sequenciado e 99,8% dos genes foram identificados e alocados nos cromossomos referência (SCHNABLE *et al.*, 2009). Além disso, os autores constataram que cerca de 85% desse genoma é composto de famílias de elementos de transposição. Essas informações foram fundamentais para o melhoramento do milho e compreensão do processo de domesticação desse grão tão importante. O genoma da variedade de milho W22 por sua vez, foi sequenciado e montado *de novo* por meio das tecnologias de sequenciamento de leitura curta (SPRINGER *et al.*, 2018). Com isso, os autores puderam constatar diferenças expressivas na estrutura genômica do W22 em comparação ao B73, com relação a diferenças na composição de transposons até polimorfismos de nucleotídeo único. A variedade Mo17 de milho, por sua vez, teve o genoma sequenciado por PacBio HiFi e Nanopore (*telomere-to-telomere* - T2T) pela leitura de moléculas ultralongas. Com 2.178,6 Mb de tamanho, a precisão foi de mais de 99,99%, revelando características estruturais de todas as regiões repetitivas do genoma (CHEN *et al.*, 2023). Com essa riqueza de dados, foi possível explorar a genômica funcional e a evolução desses genomas, o que representa um grande passo para a compressão da complexidade de diversos grupos vegetais.

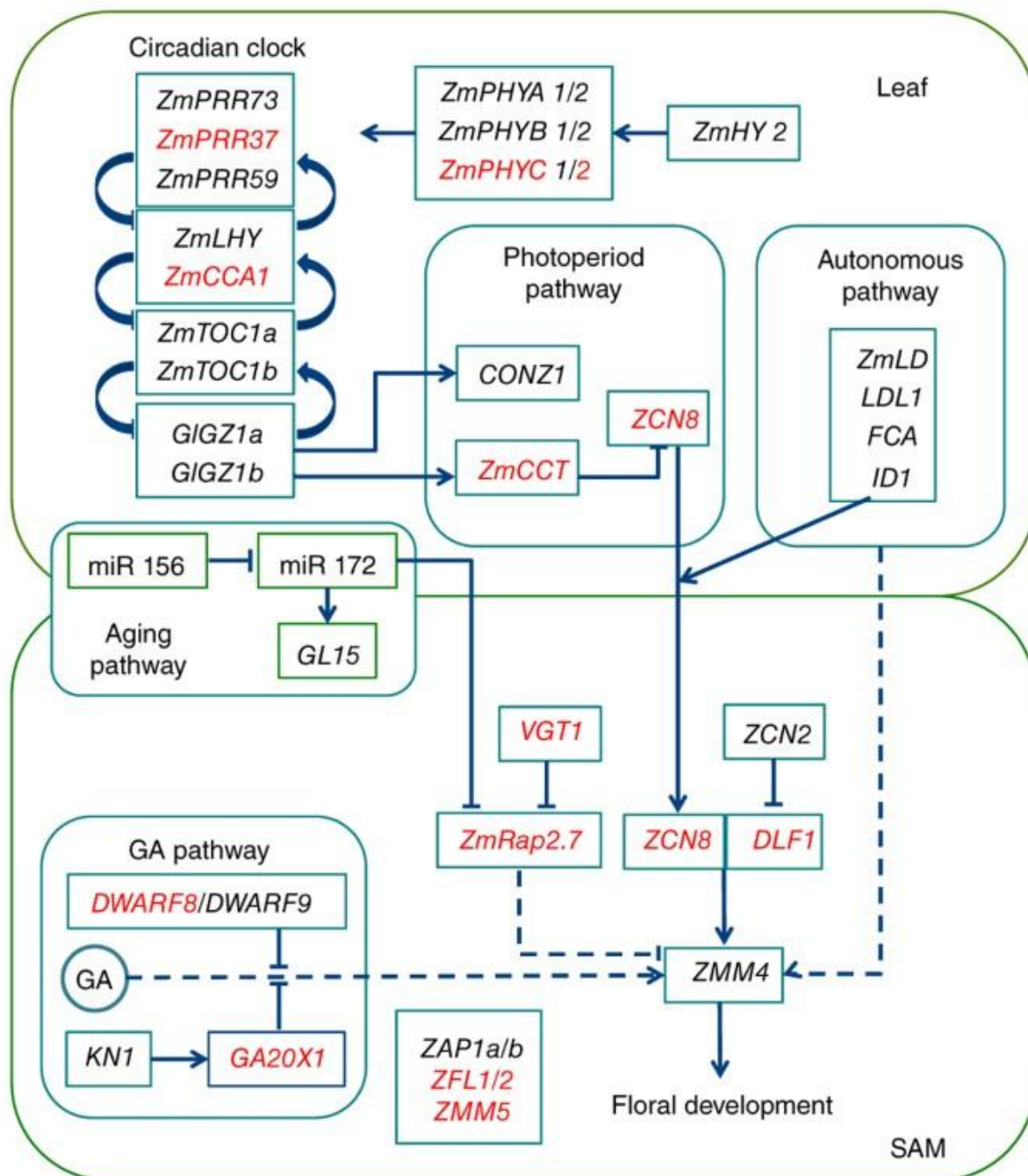
## 2.7 O TEMPO DE FLORESCIMENTO DO MILHO E SUA RELAÇÃO COM OS DNA SATÉLITES

O tempo de florescimento é uma resposta adaptativa de regulação das fases de crescimento vegetativo e reprodutivo, em relação aos efeitos bióticos, critério importante para o melhoramento do milho (GAUDINIER e BLACKMAN, 2020). O tempo de crescimento após o plantio é, em média, de nove a 10 semanas, período que corresponde à germinação, crescimento e formação do pendão. Após esse período, são necessários mais 55 dias para o florescimento e produção de grãos maduros ([www.scherer.ind.br](http://www.scherer.ind.br)). O tempo de florescimento do milho tem sido associado a variações no tamanho do genoma e à presença dos K180, que muitas vezes organizam os *knobs* no genoma do milho. JIAN *et al.* (2017) reportaram que germoplasmas com tempo de florescimento menor têm tamanho de genoma também menor, com uma baixa quantidade de K180. Por sua vez, CARVALHO *et al.* (2022) identificaram um K180 no braço curto do cromossomo 9 e, a partir de estudos de associação, sugeriram que sua presença poderia afetar o tempo de floração, tornando a cultivar mais precoce. Essas informações abriram margem para avaliarmos a relação de outras regiões com K180 próximos a outros genes, considerando as rotas metabólicas do florescimento, mesmo que em outros

cromossomos.

O florescimento é uma característica complexa, que é controlada por uma variedade de *loci* (BUCKLER *et al.*, 2009). O gene *ZCN8* (*Zea mays CENTRORADIALIS 8*), por exemplo, é um gene de florescimento, que está relacionado com a precocidade (JAFARI *et al.*, 2024). Além disso, em linhagens de milho tropicais, sensíveis ao fotoperíodo, os transcritos de *ZCN8* são regulados positivamente durante o período diurno em dias mais curtos (MENG *et al.*, 2011). Por outro lado, a expressão de *ZCN8* é regulada negativamente em períodos diurnos por *ZmCCT9* (*CONSTANS, CONSTANS-LIKE, TOC1*). Esse gene, relacionado ao tempo de florescimento do milho, possui uma inserção do transposon do tipo *Harbinger*. Esse TE está posicionado a 57 kb *upstream* de *ZmCCT9* (*CO CO-LIKE TIMING OF CAB 1 protein domain 2*), reprimindo a expressão gênica e promovendo a floração em dias longos (HUANG *et al.*, 2018). Isso ajuda a explicar como o milho se adaptou a latitudes acima de 90° (HUANG *et al.*, 2018; GAUDINIER e BLACKMAN, 2020). Outro exemplo de inserção de TEs, é o MITE (*miniature inverted-repeat transposable element*), na QTL (*Quantitative Trait Loci*) *Vgt1* (*Vegetative to generative transition1*). Esse é um elemento *cis*-regulador a cerca de 70kb *upstream* do repressor do tempo de florescimento *ZmRap2.7*, e está fortemente associado à variação do tempo de florescimento do milho (SALVI *et al.*, 2007). De acordo com ROMERO NAVARRO *et al.* (2017), há genes funcionais nos meristemas apicais relacionados ao desenvolvimento das folhas e brotos, que podem estar envolvidos no tempo de floração (Figura 2). Além disso, há uma associação de cerca de 10 e 12 genes candidatos relacionados à floração pistilada e estaminada, respectivamente, que podem estar envolvidos nas rotas metabólicas do ciclo circadiano, fotoperíodo e ácido giberélico (ROMERO NAVARRO *et al.*, 2017).

**Figura 4.** Rota metabólica dos genes envolvidos no tempo de floração no meristema apical da folha e do broto.



**Fonte:** ROMERO NAVARRO *et al.* (2017). miR: microRNA; GA: ácido giberélico.

## 2.8 CONTEÚDO DO GENOMA DO MILHO

A citometria de fluxo é muito utilizada como método de quantificação, especialmente para estimativas do conteúdo de DNA nuclear. Quando os núcleos são corados com substâncias fluorescentes e comparados com um padrão, é possível estimar a quantidade de DNA em

picogramas (pg), em que 1 pg = 978 Mbp (TEMSCH *et al.*, 2021; SLIWINSKA *et al.*, 2022). Em plantas, a aplicação dessa metodologia é muito útil para compreender como os genomas divergem entre espécies relacionadas e mostram variações da ploidia, sejam entre acessos, populações, possíveis híbridos, dentre outros exemplos (FOMICHEVA e DOMBLIDES, 2023).

O tamanho do genoma tem sido cada vez mais explorado, sendo relacionado com uma diversidade de características fenotípicas e ecológicas (SIMONIN e RODDY, 2018; RODDY *et al.*, 2020). Um estudo realizado com leguminosas do gênero *Caesalpinia* mostrou uma correlação positiva entre o tamanho dos genomas e a latitude de ocorrência das espécies (SOUZA *et al.*, 2019). Neste caso, o ambiente parece ter um papel importante para a diversidade do tamanho dos genomas nesse grupo de plantas. Similarmente, no milho, além da ampla variação fenotípica e genética, foram observadas diferenças nos tamanhos dos genomas de milhos tropicais e temperados (SPRINGER *et al.*, 2009). Contudo, estudos mais recentes no gênero *Zea* mostraram que tais variações estão relacionadas à quantidade de K180 e outras sequências satélite, além dos RT-LTRs, como o elemento da família Grande (REALINI *et al.*, 2016; SILVA *et al.*, 2020; GONZÁLEZ *et al.*, 2022). FOURASTIÉ *et al.* (2018) reportaram que nas variedades de milho crioulas da Argentina, adaptadas para altitudes entre 910 e 3.900 m acima do nível do mar, a variação nos valores 2C de DNA foi de 4,5 a 6,3 pg. Essa variação na quantidade de DNA nuclear pode implicar em variações fenotípicas, tais como taxa de crescimento, tempo de florescimento, peso das sementes e tamanho das folhas (DÍEZ *et al.*, 2013).

Análises genômicas de populações de milho com elevado valor genético são estratégicas para validar materiais específicos alvo de melhoramento. Os milhos híbridos “especiais”, em que estão inclusos os acessos crioulos e compostos, são muito importantes como fonte de diversidade para a cadeia produtiva (CECARELLI *et al.*, 1994). O banco de germoplasma da UEL tem uma excelente coleção de cultivares de milhos compostos, muitas das quais direcionadas às condições ambientais específicas da região Norte do Paraná, bem como de outras regiões do país. Tais cultivares crioulas de milho branco, doce e super doce, e com diferenças no tempo de florescimento, ainda não foram avaliadas do ponto de vista citogenético e genômico. Por isso, há margem para explorar essa fonte de amostras em busca de marcas cromossômicas úteis para os programas de melhoramento do milho.

### 3 OBJETIVO GERAL E ESPECÍFICOS

#### 3.1 OBJETIVO GERAL

Avaliar a diversidade de sequências repetitivas de DNA em diferentes cariótipos de cultivares de milhos compostos, utilizando rastreamento dessas sequências por FISH e por ferramentas de bioinformática, comparando-a às linhagens de referência NAM, em relação ao tempo de florescimento.

#### 3.2 OBJETIVOS ESPECÍFICOS

1. Identificar heteromorfismos cromossômicos utilizando sondas de DNA repetitivo dos tipos K180, TR-1, CentC e DNAr 5S, em milhos compostos;
2. Relacionar o tempo de florescimento dos milhos compostos com a abundância dos sítios de K180 e TR-1;
3. Mensurar os valores C de DNA dos milhos compostos e relacioná-los com o tempo de florescimento;
4. Contrastar o tempo de florescimento das linhagens de milho NAM com a distribuição dos sítios de K180, TR-1, CentC e com a proximidade desses aos genes do florescimento nos pseudocromossomos;
5. Caracterizar os TEs inseridos ou acerca dos genes de florescimento, tendo como referência os pseudocromossomos das linhagens de milho NAM;

#### 4 BIBLIOGRAFIA

- ALBERT, P. S. et al. Whole-chromosome paints in maize reveal rearrangements, nuclear domains, and chromosomal relationships. **Proceedings of the National Academy of Sciences**, v. 116, n. 5, p. 1679-1685, 2019.
- ANANIEV, E. V.; PHILLIPS, R. L.; RINES, H. W. A K180-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome K180 megatransposons?. **Proceedings of the National Academy of Sciences**, v. 95, n. 18, p. 10785-10790, 1998.
- ANANIEV, E. V.; PHILLIPS, R. L.; RINES, H. W. Complex structure of knob DNA on maize chromosome 9: retrotransposon invasion into heterochromatin. **Genetics**, v. 149, n. 4, p. 2025-2037, 1998.
- ANDRADE, L. R. B. de et al. Genetic vulnerability and the relationship of commercial germplasms of maize in Brazil with the nested association mapping parents. **PLoS One**, v. 11, n. 10, p. e0163739, 2016.
- BRAZ, G. T. et al. A universal chromosome identification system for maize and wild *Zea* species. **Chromosome Research**, v. 28, p. 183-194, 2020.
- BOREM, A.; MIRANDA, G.V.; FRITSCHÉ-NETO, R. **Melhoramento de plantas**, 6<sup>a</sup> ed. Viçosa: Editora UFV. 2013.
- BUCKLER, E. S. et al. The genetic architecture of maize flowering time. **Science**, v. 325, n. 5941, p. 714-718, 2009.
- CARVALHO, R. F. et al. A heterochromatic knob reducing the flowering time in maize. **Frontiers in Genetics**, v. 12, p. 799681, 2022.
- CECCARELLI, S. Specific adaptation and breeding for marginal conditions. **Euphytica**, v. 77, n. 3, p. 205-219, 1994.
- CUI, L. et al. Karyotype analysis, genomic and fluorescence in situ hybridization (GISH and FISH) reveal the ploidy and parental origin of chromosomes in *Paeonia itoh* hybrids. **International Journal of Molecular Sciences**, v. 23, n. 19, p. 11406, 2022.
- CHEN, J. et al. A complete telomere-to-telomere assembly of the maize genome. **Nature genetics**, v. 55, n. 7, p. 1221-1231, 2023.
- CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. Acompanhamento da safra brasileira de grãos. Safra 2023/24 – 12º Levantamento. Brasília: Companhia Nacional de Abastecimento. 2024. Disponível em: <https://www.conab.gov.br/info-agro/safra/graos/boletim-da-safra-de-graos>. Acesso em: 20 set. 2024.
- COSTA, F. M.; SILVA, N. C. A.; OGLIARI, J. B. Maize diversity in southern Brazil: indication of a microcenter of *Zea mays* L. **Genetic resources and crop evolution**, v. 64, p. 681-700, 2017.
- DENNIS, E. S.; PEACOCK, W. J. Knob heterochromatin homology in maize and its relatives. **Journal of molecular evolution**, v. 20, p. 341-350, 1984.
- DÍEZ, C. M. et al. Genome size variation in wild and cultivated maize along altitudinal gradients. **New Phytologist**, v. 199, n. 1, p. 264–276, 2013.

- EICKBUSH, T. H.; MALIK, H. S. Origins and evolution of retrotransposons. **Mobile DNA ii**, p. 1111-1144, 2007.
- EICHOLZ, E. D. et al. Milhos no cadastro nacional de variedades locais ou crioulos para o Rio Grande do Sul. Pelotas: Embrapa Clima Temperado, p. 35, 2018.
- FAOSTAT. Food and Agriculture Organization of the United Nations. Crops and livestock products. Roma: Food and Agriculture Organization of the United Nations. 2024. Disponível em: <https://www.fao.org/faostat/en/#data/QCL>. Acesso em: 05 set. 2024.
- FAROOQ, A. et al. A critical review of climate change impact at a global scale on cereal crop production. **Agronomy**, v. 13, n. 1, p. 162, 2023.
- FESCHOTTE, C. Transposable elements: McClintock's legacy revisited. **Nature Reviews Genetics**, v. 24, n. 11, p. 797-800, 2023.
- FOMICHEVA, M.; DOMBLIDES, E. Mastering DNA content estimation by flow cytometry as an efficient tool for plant breeding and biodiversity research. **Methods and Protocols**, v. 6, n. 1, p. 18, 2023.
- FOURASTIÉ, M. F. et al. Are cytological parameters of maize landraces (*Zea mays* ssp. *mays*) adapted along an altitudinal cline?. **Journal of plant research**, v. 131, p. 285-296, 2018.
- GARCÍA-LARA, S.; SERNA-SALDIVAR, S. O. Corn history and culture. **Corn**, p. 1-18, 2019.
- GAUDINIER, A.; BLACKMAN, B. K. Evolutionary processes from the perspective of flowering time diversity. **New Phytologist**, v. 225, n. 5, p. 1883-1898, 2020.
- GENT, J. I. et al. Genomics of maize centromeres. **The Maize Genome**, p. 59-80, 2018.
- GOMEZ-CASATI, D. F. et al. Applications of bioinformatics to plant biotechnology. **Current issues in molecular biology**, v. 27, n. 1, p. 89-104, 2018.
- GONZÁLEZ, G. E. et al. Causes and consequences of DNA content variation in *Zea*. **Journal of the Argentine Society of Genetics**, v. 33, n. 1 p. 43-49, 2022.
- GUZZON, F. et al. Conservation and use of Latin American maize diversity: Pillar of nutrition security and cultural heritage of humanity. **Agronomy**, v. 11, n. 1, p. 172, 2021.
- HABERER, G. et al. European maize genomes highlight intraspecies variation in repeat and gene content. **Nature Genetics**, v. 52, n. 9, p. 950-957, 2020.
- HAN, H. et al. Chromosome-level genome assembly of *Solanum pimpinellifolium*. **Scientific Data**, v. 11, n. 1, p. 577, 2024.
- HASSAN, A. H.; MOKHTAR, M. M.; EL ALLALI, A. Transposable elements: multifunctional players in the plant genome. **Frontiers in Plant Science**, v. 14, p. 1330127, 2024.
- HEUBERGER, M. et al. Evolution of Einkorn wheat centromeres is driven by the mutualistic interplay of two LTR retrotransposons. **Mobile DNA**, v. 15, n. 1, p. 16, 2024.
- HOUBEN, A. et al. CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. **Chromosoma**, v. 116, p. 275-283, 2007.
- HUANG, C. et al. *ZmCCT9* enhances maize adaptation to higher latitudes. **Proceedings of the National Academy of Sciences**, v. 115, n. 2, p. E334-E341, 2018.

- JAFARI, F. et al. Breeding maize of ideal plant architecture for high-density planting tolerance through modulating shade avoidance response and beyond. **Journal of Integrative Plant Biology**, v. 66, n. 5, p. 849-864, 2024.
- JIAN, Y. et al. Maize (*Zea mays* L.) genome size indicated by 180-bp knob abundance is associated with flowering time. **Scientific Reports**, v. 7, n. 1, p. 5954, 2017.
- JIANG, J.; GILL, B. S. Current status and the future of fluorescence *in situ* hybridization (FISH) in plant genome research. **Genome**, v. 49, n. 9, p. 1057-1068, 2006.
- JIANG, J. Fluorescence in situ hybridization in plants: recent developments and future applications. **Chromosome Research**, v. 27, n. 3, p. 153-165, 2019.
- JIN, W. et al. Maize centromeres: organization and functional adaptation in the genetic background of oat. **The Plant Cell**, v. 16, n. 3, p. 571-581, 2004.
- KATO, A.; LAMB, J. C.; BIRCHLER, J. A. Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. **Proceedings of the National Academy of Sciences**, v. 101, n. 37, p. 13554-13559, 2004.
- KIROV, I. et al. Functional *Allium fistulosum* centromeres comprise arrays of a long satellite repeat, insertions of retrotransposons and chloroplast DNA. **Frontiers in Plant Science**, v. 11, p. 562001, 2020.
- KISTLER, L. et al. Multiproxy evidence highlights a complex evolutionary legacy of maize in South America. **Science**, v. 362, n. 6420, p. 1309-1313, 2018.
- KOLAWOLE, A. O. et al. Genetic advance in grain yield and other traits in two tropical maize composites developed via reciprocal recurrent selection. **Crop Science**, v. 58, n. 6, p. 2360-2369, 2018.
- LI, W. et al. Chromosome genome assembly and annotation of Adzuki Bean (*Vigna angularis*). **Scientific Data**, v. 11, n. 1, p. 1074, 2024.
- LONGLEY, A. E. Chromosome morphology in maize and its relatives. **Botanical Review**, v. 7, n. 5, p. 263-289, 1941.
- MCCLINTOCK, B. Chromosome morphology in *Zea mays*. **Science**, v. 69, n. 1798, p. 629-629, 1929.
- MCMULLEN, M. D. et al. Genetic properties of the maize nested association mapping population. **Science**, v. 325, n. 5941, p. 737-740, 2009.
- MENG, X.; MUSZYNSKI, M. G.; DANILEVSKAYA, O. N. The FT-like *ZCN8* gene functions as a floral activator and is involved in photoperiod sensitivity in maize. **The Plant Cell**, v. 23, n. 3, p. 942-960, 2011.
- MIKEL, M. A.; DUDLEY, J. W. Evolution of North American dent corn from public to proprietary germplasm. **Crop science**, v. 46, n. 3, p. 1193-1205, 2006.
- NEUMANN, P. et al. Plant centromeric retrotransposons: a structural and cytogenetic perspective. **Mobile DNA**, v. 2, p. 1-16, 2011. PANZEA. Earlier PANZEA. New York. PANZEA. 2024. Disponível em: <https://www.panzea.org/history-of-panzea>. Acesso em: 30 out. 2024.
- PERALES, H.; GOLICHER, D. Mapping the diversity of maize races in Mexico. **PLoS one**, v. 9, n. 12, p. e114657, 2014.

- REALINI, M. F. et al. Intra-specific variation in genome size in maize: cytological and phenotypic correlates. **AoB Plants**, v. 8, p. plv138, 2016.
- REALINI, M. F. et al. Exploring karyotype diversity of Argentinian Guaraní maize landraces: relationship among South American maize. **PLoS One**, v. 13, n. 6, p. e0198398, 2018.
- REVILLA, P. et al. Traditional foods from maize (*Zea mays* L.) in Europe. **Frontiers in Nutrition**, v. 8, p. 683399, 2022.
- RODDY, A. B. et al. The Scaling of Genome Size and Cell Size Limits Maximum Rates of Photosynthesis with Implications for Ecological Strategies. **International Journal of Plant Sciences**, v. 181, n. 1, p. 75–87, 2020.
- ROMAY, M. C. et al. Three cycles of full-sib reciprocal recurrent selection in two Spanish maize populations. **Crop Science**, v. 51, n. 3, p. 1016-1022, 2011.
- ROMERO NAVARRO, J. A. et al. A study of allelic diversity underlying flowering-time adaptation in maize landraces. **Nature genetics**, v. 49, n. 3, p. 476-480, 2017.
- SAHU, S. K.; LIU, H. Long-read sequencing (method of the year 2022): the way forward for plant omics research. **Molecular Plant**, v. 16, n. 5, p. 791-793, 2023.
- SARAIVA, E. A. et al. Desempenho agrônômico de variedades de milhos crioulos cultivados no Vale do Jequitinhonha. **Revista Brasileira de Agropecuária Sustentável**, v. 9, n. 2, p. 33-43, 2019.
- SCHERER. Plantio de Milho: Em Quanto Tempo Ocorre o Florescimento. Paraná: SCHERER. 2019. Disponível em: <https://www.scherer.ind.br/noticias/detalhe/plantio-de-milho-em-quanto-tempo-ocorre-o-florescimento#:~:text=O%20tempo%20de%20crescimento%20e,e%20produ%C3%A7%C3%A3o%20dos%20gr%C3%A3os%20maduros>. Acesso em: 15 out. 2024.
- SCHNABLE, P. S. et al. The B73 maize genome: complexity, diversity, and dynamics. **science**, v. 326, n. 5956, p. 1112-1115, 2009.
- SCHNEIDER, K. L. et al. Inbreeding drives maize centromere evolution. **Proceedings of the National Academy of Sciences**, v. 113, n. 8, p. E987-E996, 2016.
- SERRATOS, J. A. The origin and diversity of maize in the American continent. **Mexico: Greenpeace México**, 2009.
- SILVA, J. C. et al. Repetitive sequences and structural chromosome alterations promote intraspecific variations in *Zea mays* L. karyotype. **Scientific reports**, v. 10, n. 1, p. 8866, 2020.
- SIMONIN, K. A.; RODDY, A. B. Genome downsizing, physiological novelty, and the global dominance of flowering plants. **PLOS Biology**, v. 16, n. 1, p. e2003706, 2018.
- SLIWINSKA, E. et al. Application-based guidelines for best practices in plant flow cytometry. **Cytometry Part A**, v. 101, n. 9, p. 749-781, 2022.
- SOUZA, G. et al. Do tropical plants have smaller genomes? Correlation between genome size and climatic variables in the *Caesalpinia* Group (Caesalpinioideae, Leguminosae). **Perspectives in plant ecology, evolution and systematics**, v. 38, p. 13-23, 2019.
- SPRINGER, N. M. et al. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content. **PLoS genetics**, v. 5, n. 11, p.

e1000734, 2009.

SPRINGER, N. M. et al. The maize W22 genome provides a foundation for functional genomics and transposon biology. **Nature genetics**, v. 50, n. 9, p. 1282-1288, 2018.

STACK, S. M. et al. Meiotic crossing over in maize knob heterochromatin. **Genetics**, v. 205, n. 3, p. 1101-1112, 2017.

STITZER, M. C. et al. The genomic ecosystem of transposable elements in maize. **PLoS Genetics**, v. 17, n. 10, p. e1009768, 2021.

TAN, Y. C. et al. Bioinformatics approaches and applications in plant biotechnology. **Journal of Genetic Engineering and Biotechnology**, v. 20, n. 1, p. 106, 2022.

THAKUR, J.; PACKIARAJ, J.; HENIKOFF, S. Sequence, chromatin and evolution of satellite DNA. **International journal of molecular sciences**, v. 22, n. 9, p. 4309, 2021.

USDA. U. S. Department of Agriculture. Production - Corn. Washington: U. S. Department of Agriculture. 2024. Disponível em: <https://fas.usda.gov/data/production/commodity/0440000>. Acesso em: 28 set. 2024.

WANG, J. et al. Effect of climate change on the yield of cereal crops: A review. **Climate**, v. 6, n. 2, p. 41, 2018.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature reviews genetics**, v. 8, n. 12, p. 973-982, 2007.

WIDARMI, W. D. et al. Characterization of lettuce chromosomes based on condensation patterns and physical mapping of 45S and 5S rDNAs using fluorescence in situ hybridization. **Cytologia**, v. 85, n. 1, p. 49-55, 2020.

WOODHOUSE, M. R. et al. A pan-genomic approach to genome databases using maize as a model system. **BMC plant biology**, v. 21, p. 1-10, 2021.

ZHANG, L. et al. Identification of peanut (*Arachis hypogaea*) chromosomes using a fluorescence in situ hybridization system reveals multiple hybridization events during tetraploid peanut formation. **New Phytologist**, v. 211, n. 4, p. 1424-1439, 2016.

ZHONG, C. X. et al. Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. **The Plant Cell**, v. 14, n. 11, p. 2825-2836, 2002.

ZHOU, J. et al. Centromeres: From chromosome biology to biotechnology applications and synthetic genomes in plants. **Plant Biotechnology Journal**, v. 20, n. 11, p. 2051-2063, 2022.

## 5 ARTIGO

O artigo será submetido a revista Genome Research.

### **Insights into the relationships between the distribution of repetitive DNA sequences and the flowering time of maize**

Juliana Machado da Silva<sup>a,b</sup>, Rafaela Rodrigues Pinheiro<sup>a</sup>, Thamirys Silva Valentim<sup>a,b</sup>, Josué Maldonado Ferreira<sup>c</sup>, Mateus Mondin<sup>d,\*</sup>, André Luís Laforga Vanzela<sup>a,b,\*</sup>

<sup>a</sup>Laboratório de Citogenética e Diversidade Vegetal (LCDV), Departamento de Biologia Geral, CCB, Universidade Estadual de Londrina, Londrina, 86097-570, Paraná, Brazil.

<sup>b</sup>Programa de Pós-graduação em Genética e Biologia Molecular, CCB, Universidade Estadual de Londrina, Londrina, 86097-570, Paraná, Brazil.

<sup>c</sup>Laboratório de Melhoramento Genético (LMG), Departamento de Biologia Geral, CCB, Universidade Estadual de Londrina, Londrina, 86097-570, Paraná, Brazil.

<sup>d</sup>Laboratório de Citogenômica e Epigenética, Departamento de Genética da Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ), Piracicaba, 13418-900, São Paulo, Brazil.

\*Corresponding authors: [andrevanzela@uel.br](mailto:andrevanzela@uel.br) and [mmondin@usp.br](mailto:mmondin@usp.br)

da Silva J.M. - ORCID:0000-0003-3387-2449

Pinheiro R.R. - ORCID:0000-0001-8270-464X

Valentim T.S. - ORCID:0009-0000-7968-9530

Ferreira J.M. - ORCID:0000-0002-0085-6891

Mondin M. - ORCID:0000-0002-7641-0546

Vanzela A.L.L. - ORCID: 0000-0002-2442-2211

#### **Abstract**

*Zea mays* is an ideal model for cytogenomic studies due to the high proportion of repetitive DNA families in its genome. The most prominent are the K180 and TR-1 families. Both can form heterochromatin blocks and influence genetic recombination and expression of nearby genes. Flowering time is a crucial criterion in plant breeding. This period seems to be associated with variations in genome size and the presence of K180. Another significant component of repetitive fraction is transposable elements (TEs), which constitute approximately 85% of the maize genome. These elements have reshaped the genomic landscape by altering the arrangement of other sequences. This study aimed to evaluate the karyotypic diversity in composite maize varieties through the screening of repetitive DNA sequences by fluorescence *in situ* hybridization (FISH) and bioinformatics. The analyses were performed in a comparative framework, using the NAM lines as a reference point. For this purpose, we used DNAsats (K180, TR-1 and CentC) and 5S DNA probes for FISH, along the estimation of 2C DNA values. In the NAM lines, pseudochromosomes were mapped bioinformatically to 14 key genes in the flowering pathway. The composite maize karyotypes showed heteromorphisms and polymorphisms in FISH signals. In contrast, the 2C DNA values remained relatively consistent. Pseudochromosome analysis revealed considerable variability in both the number and size of K180 and TR-1 arrays. The arrangements of smaller arrays showed the greatest diversity in position and size. An analysis of the DNAsat arrays in relation to their distance from flowering

genes indicated a scarcity of monomers and a considerable distance from the target genes. Regarding the frequency of retrotransposons near genes associated with flowering time, we observed considerable diversity into gene-to-gene insertions. In general, two genes showed a remarkable prevalence of RT-LTR insertions: *zfl2* and *mads1*. Overall, our findings indicate that the number, size, and location of the DNAsats (K180 and TR-1) along with genome size were insufficient to establish a correlation with flowering time in composite maize.

### **Keywords**

Constitutive heterochromatin; Flowering time; Genome; K180; Maize; Transposable elements

### **Introduction**

*Zea mays* L. (Poaceae) evolved approximately 9,000 years ago from its wild teosinte ancestor, *Z. mays* ssp. *parviglumis*, with introgression from *Zea mays* ssp. *mexicana*, where it was domesticated and diversified (Piperno et al. 2001; Kistler et al. 2018; Yang et al. 2023). Maize is an excellent model for plant cytogenetics, as its chromosomes in the pachytene stage exhibit cytological characteristics that allow their identification. Additionally, the presence of knobs and their variability, particularly regarding the possibility of appearing in heterozygous configurations, further enhance its significance in cytogenetic studies. The heterochromatic knobs in maize can be composed of two distinct families of satellite DNA: K180 (180 bp length) and TR-1, with 350 bp (Dennis and Peacock 1984; Ananiev et al. 1998a). Together, these sequences constitute about 20% of tandem repetitive sequences in the maize genome (Hufford et al. 2021; Chen et al. 2023). The arrangement of these satellites is quite variable in maize traditional varieties, inbred lines, hybrids and modern cultivars, ranging from small interstitial blocks to large arrays located predominantly in the distal chromosomal regions (McClintock et al. 1981; Aguiar-Perecin et al. 2000; Kato et al. 2004; Mondin et al. 2014). Heterochromatic knobs are found in gene-rich regions and have been shown to interfere with recombination patterns. Furthermore, they can exhibit neocentromeric activity and are believed to affect phenotypic traits, such as flowering time (Rhoades et al. 1942; Ghaffari et al. 2013; Kreiner and Wright 2018; Carvalho et al. 2020; Swentowsky et al. 2020). The proximal chromosomal regions of maize are also attractive because they are rich in a conserved repetitive DNA known

as CentC (centromere repeat C) with a length of 156 bp, interspersed with centromeric retrotransposons (CRMs), a gypsy lineage (Ananiev et al. 1998b).

Another important fraction of repetitive sequences consists of transposable elements (TEs) make up the majority of plant genomes (Wicker et al. 2007; Hassan et al. 2024). They make up approximately 85% of the maize genome, with retrotransposons being the most abundant class (Schnable et al. 2009; Stitzer et al. 2021). These elements can transpose or copy themselves within the genome using self-encoded proteins. Depending on their insertion and accumulation, they can cause DNA breakage, promote unequal recombination, create or revert mutations, alter gene expression, genotypes, and phenotypes, and can be the main driver of genome evolution, as observed in maize mutants with variegated grains (Zattera and Bruschi 2022; Feschotte 2023, Hassan et al. 2024).

The chromosome number in maize is  $2n = 20$ , except for the presence of B chromosomes (Longley 1941; Birchler and Yang, 2021). However, the karyotypes of different maize varieties exhibit significant variation due to the dynamics of the repetitive fraction in the genome. Detailed chromosome mapping using different probes from the repetitive fraction, as well as by the use of oligo-FISH probes, improved our understanding on the reorganization of karyotypes in different maize varieties, especially by the difficulty of accurately identifying some homologous pairs (Kato et al. 2004; Albert et al. 2018; Jiang et al. 2019; Braz et al. 2020).

The most widely used probes for chromosome mapping in maize include K180, TR-1, CentC, Cent4, subtelomeric, and B-specific repetitive DNA. K180 and TR-1 probes have been used to understand the intra- and interpopulation diversity among Argentine maize landraces (Realini et al. 2018). In some cases, these probes have been used to associate variations in satellite DNA (satDNA) fraction with potential upland cultivation (Fourastié et al. 2018). The accumulation and retraction of repetitive fraction are generally associated with fluctuations in genomic DNA content. Variations in C DNA values have been correlated with

a wide range of phenotypic and ecological traits (Souza et al. 2022; Roddy et al. 2020). Studies in maize have shown differences in genome size between varieties adapted to tropical and temperate regions (Rayburn et al. 1990; McMurphy and Rayburn, 1991; Poggio et al. 2013; Díez et al. 2013; Realini et al. 2016). Fourastié et al. (2018) reported a pronounced variation in 2C DNA values (ranging from 4.5 to 6.3 pg) in Argentine maize landraces, which are adapted to altitudes between 910 and 3,900 m above sea level. Additionally, variations in DNA C-values have been associated with differences in the amount of satDNA (K180) and LTR retrotransposons (Realini et al. 2016; Bilinski et al. 2018; Silva et al. 2020; González et al. 2022). These variations may influence key phenotypic traits, such as growth rate, flowering time, seed weight and leaf size (Díez et al. 2013). Jian et al. (2017) suggested that the reduction in the number of K180 and the decrease in genome size could be associated with the earlier flowering time in maize. The flowering precocity may be related to the K180 location on the short arm of chromosome 9 (Carvalho et al. 2022). However, these analyses did not consider the small arrangements of satDNA monomers (K180 and TR-1) near genes related to flowering time or estimate the distance between them. The same limitation applies to other families of repetitive DNA, such as retrotransposons.

Maize breeding programs aim to produce varieties with desirable agronomic traits and enhanced productivity. The production of composite maize varieties through crossbreeding with multiple parental lines results in the generation of new varieties with high genetic diversity (Borem et al. 2013; Kolawole et al. 2018). Although this is a common practice, few cytogenomic studies have focused on the karyotypic diversity of composite maize derived from landrace parents. In this context, a comprehensive comparative analysis of maize landrace varieties of composite origin is crucial for understanding their karyotype dynamics. We measured the DNA C-values and conducted comparative mapping using different satDNA sequences (K180, TR-1 and CentC) and 5S rDNA as FISH probes in different composite maize

varieties. Additionally, we performed bioinformatic analysis to track repetitive sequences and genes in the pseudochromosomes of lineage panel Nested Association Mapping (NAM) (Hufford et al. 2021), enabling us to access the proximity of satDNA arrays to target genes. Here, we explore the dynamics of the satDNA accumulation and their relationship with the flowering time in maize.

## **Material and Methods**

### *Plant Materials*

The six varieties of composite maize selected for the study were obtained after the first cross to maintain greater similarity to the parents. These materials are part of the germplasm bank of the Genetic Improvement Laboratory (LMG), Department of General Biology, State University of Londrina (Londrina, Paraná, Brazil.) They were selected for grain quality (white or yellow), cultivation time and flowering time (Supplementary Table 1). Reference B73 was provided by the Laboratory of Cytogenomics and Epigenetics, Department of Genetics, Luiz de Queiroz Agricultural School (ESALQ), Piracicaba, São Paulo, Brazil. The samples were grown in a greenhouse to obtain roots for cytogenomic analyses and leaves for extraction of genomic DNA and nuclei for estimation of C DNA values by flow cytometry.

### *Fluorescence in situ hybridization (FISH)*

The 5S rDNA probe was obtained from PCR using template DNA from B73 cultivar, extracted according to Doyle and Doyle (1990), with minor modifications. The 120 bp length conserved 5S rDNA coding region was amplified with PCR using the primers UP46-F (5' GTGCGATCATAACCAGCRYTAATGCACCGG) and UP47-R (5' GAGGTGCAACACGAGGACT TCCCAGGAGG). PCR was conducted using a mix containing 2 mM MgCl<sub>2</sub>, 0.4 μM of each primer, 0.2 mM dNTP, ~30 ng of template, 1.25 U

of Taq polymerase and ultrapure water in a final volume of 25  $\mu$ L. Probes were labeled using 0.2 mM dNTP, containing dGTP (25%), dCTP (25%), dATP (25%), dTTP (17.5%) and Cy3-dUTP (7.5%). PCR was performed under the following conditions: 94 °C for 2 min, 30 cycles of 94 °C for 40 s, 62 °C for 50 s and 72 °C for 1 min and a final extension of 72 °C for 10 min. The probe for TR-1 satDNA (350 bp length) was obtained by PCR but using the pGEM-T vector carrying TR-1 as a template. The M13 primer set (F-CTGGCCGTCGTTTAC and R-CAGGAAACAGCTATGAC) flanking the insert was used with the following reaction conditions: 3 min at 94 °C, 34 cycles of 1 min at 94 °C, 1 min at 58 °C and 1 min at 72 °C, followed by 5 min at 72 °C. The PCR product was verified on a 1% agarose gel and the probe was labeled with biotin-16-dUTP using the nick translation kit (Jena Bioscience). The K180 and CentC probes were commercially purchased as oligomers labeled with 6-FAM (K180 with 93 bp) and TAMRA (CentC with 78 bp;) from Exxtend Company (<https://www.exxtend.com.br/>).

Root tips pretreated with 2 mM 8-hydroxyquinoline at 28 °C (2 h) and fixed in ethanol-acetic acid (3:1, v:v) solution. Samples were treated in a solution composed of 2% cellulase (0.3 units/mg) Onozuka R-10 (Duchefa Biochemie) and 20% pectinase (5 units/mg - Sigma) at 37 °C (2 h) and squashed in a drop of 60% acetic acid. After liquid nitrogen freezing, the coverslips were removed, and the slides were air-dried at room temperature. For FISH, the slides were denatured at 85 °C for 2 min in a mixture containing 50% formamide, 10% dextran sulphate, 2 $\times$  SSC, and 100 ng/ $\mu$ L of each probe. Post-hybridization washes were carried out with 70% stringency using 2 $\times$  SSC at 55-60 °C for 20 min. Finally, the slides were dehydrated and counterstained with DAPI in Vectashield medium. The slides were analyzed in three replicates for at least 10 cells. Chromosome images were acquired using a Leica DM 4500B microscope, equipped with a DFC300FX camera. The images were optimized for contrast and brightness using GIMP 2.8 Image Editor. Idiograms were constructed using CorelDraw, based

in the average size of chromosome arms from at least five metaphases with FISH signals were positioned. All pairs were aligned based on centromere (CentC) positioning.

#### *Estimation of DNA C-values*

The DNA content of six cultivars of composed maize, as well as the reference B73, was measured and compared according to the method of Praça-Fontes et al. (2011), using *Solanum lycopersicum* L. ‘Stupické’ ( $2C = 2.00 \text{ pg}$ ) as the standard. Younger leaves were chopped in 250  $\mu\text{L}$  of OTTO-I buffer, containing polyethylene glycol (PEG) (7%) and RNase ( $1 \text{ mg/mL}^{-1}$ ). Samples were filtered through 30  $\mu\text{m}$  nylon mesh and centrifuged at  $500 \times g$ . Nuclei were then stained with the OTTO-I: OTTO-II buffer, 1:2, v:v (Loureiro et al. 2006), containing PEG (7%), RNase ( $1 \text{ mg/mL}^{-1}$ ) and propidium iodide ( $1 \text{ mg/mL}^{-1}$ ). The samples were again filtered through 20  $\mu\text{m}$  nylon mesh. Measurements were performed using an Accuri C6 BD flow cytometer (Becton, Dickinson and Company), according to the manufacturer's instructions. Each measurement had at least 5,000 events, considering a coefficient of variation (CV)  $<4\%$  (Temsch et al. 2022). For each accession, three replicates were performed on different plants of the same cultivar on two different days, as recommended by Doležel et al. (2007), giving a total of six measurements for each cultivar. The absolute amount of DNA in a sample is calculated from the values of the average G1 peaks. GraphPad Prism V. 10.2 software ([www.graphpad.com](http://www.graphpad.com)) was used for statistical analysis. The data were subjected to the normality test (Shapiro-Wilk), followed by the parametric test of variance (one-way ANOVA). Finally, the Tukey post-test was used to indicate significant differences between treatments (B73  $\times$  C1  $\times$  C2  $\times$  C3  $\times$  C4  $\times$  C5  $\times$  C6  $\times$  C7). P values  $< 0.05$  were considered significant.

#### *Characterizing the repetitive fraction from selected maize NAM inbred lines*

Genomes of seven maize varieties from the NAM collection were mapped (Mcmuller

et al. 2009; Hufford et al. 2021). The genomes, sequenced by PacBio (Pacific Biosciences) and with end-to-end chromosome assemblies, were obtained from the NCBI database (<https://www.ncbi.nlm.nih.gov/>). The genome selection criterion was flowering time, based on the number of days to panicle emergence and the number of days to stigma emergence of the NAM population (Buckler et al. 2009). Therefore, we selected two genomes with a late flowering profile (CML245 and CML52), two with an early flowering profile (P39 and IL14H), two with an intermediate flowering profile (M37W and NC350), in addition to the reference genome B73 (version 5), which also has an intermediate flowering pattern (Supplementary Table 2).

Fourteen genes related to the flowering time metabolic pathway (Romero Navarro et al. 2017), with sequences available in the curated MaizeGDB database (<https://www.maizegdb.org>), were used (Supplementary Table 3). The sequences of the satDNA K180 monomers, 180 bp long (GenBank: M32532.1) and TR-1, 350 bp long (GenBank: AF071121.1) were obtained from NCBI and mapped to the pseudochromosomes of the seven genomes analyzed. Two different tools were used for this purpose: BLASTn (Altschul et al. 1990) and RepeatMasker (Smit et al. 2013), both with default parameters, selection criteria of >80% identity and >90% sequence size. The same was done for the 14 target genes, with a criterion of 100% identity.

Once these sequences were aligned in the pseudochromosomes of the seven NAM genomes, the values of each position were used to amplify each fragment (50,000 bases upstream and 50,000 bases downstream of each gene) to extract the flanking sequences of each gene. This procedure was performed using the extractseq tool of the EMBOSS package (Rice et al. 2000). This strategy allowed the identification of the repetitive fraction of satDNA and RT-LTRs near or inserted into the target genes. In this analysis, the size, i.e. the number of monomers of each satDNA array for the K180 and TR-1, and the distance of each array from

the flowering genes were taken into account. Only distances less than 175 kbp were considered, according to the criteria described by Voguel et al. (2009). These authors suggested that this distance would be the maximum limit at which heterochromatin could cause some physical position effect on a nearby gene. Regarding the presence of transposable elements, the occurrence, size, distance and insertion of retrotransposons (Gypsy and Copia) in the same genes were evaluated, taking into account the same distance described above. The search was based on the annotation and curation available on the MaizeGDB JBrowse platform (<https://jbrowse.maizegdb.org/>) by screening: Repetitive Elements - Transposable Elements. The graphs representing this mapping were generated using the tools SRplot (Tang et al. 2023) and Rstudio (Team 2020).

## Results

### *Chromosome mapping in composite maize using FISH*

We used a set of repetitive DNA probes (K180, TR-1, CentC, and 5S rDNA) for FISH to analyze six composite maize samples (hereafter referred to as Composite 1 to 7) and the B73 inbred line as a reference. Chromosome mapping revealed polymorphisms and heteromorphisms among the analyzed samples (Figures 1, 2, and 3), except for 5S rDNA loci, which exhibited consistent FISH signals on the chromosome 2 pair across all analyzed samples (Figure 2).

For the satDNA probes (K180 and TR-1), we observed variations in the position, size/intensity, and number of FISH signals. B73 and Composite 7 displayed the lowest number of K180 blocks (12 and 13 blocks, respectively), whereas Composites 1 and 4 exhibited the highest numbers, with 17 blocks (Figures 1A-B, D, and F). Variability in the size/intensity of the hybridization signals was common within and between karyotypes (Supplementary Table

4).

Composite maize samples displayed greater variability in TR-1 DNAsat arrays, compared to the K180 signals (Figures 2 and 3). In B73, we observed four hybridization signals with the TR-1 probe (Figure 3A). Composite 7 exhibited only two FISH signals for TR-1 in (Figure 3F), whereas Composite 2 showed nine hybridization sites (Figures 2E and 3C). These composites exhibited polymorphisms and heteromorphisms TR-1 signals on both homologues of chromosome 2 pair, whereas Composite 1 was heteromorphic, colocalizing with the 5S rDNA signal (Figures 3C, F and B, respectively). We observed TR-1 FISH signals on chromosome 6 in all composites except Composite 7 (Figure 3F). This pair also displayed heteromorphism in Composites 1 and 6 (Figures 3D and E). Additionally, weak dots were also found in one chromosome of Composite 4, but it was not possible to determine the exact location of this probe on chromosome 7 or 9 (Figure 3D). Differences in the number of monomers per TR-1 block between homologs as well as between heterologous.

FISH with the CentC probe showed consistent signals in all centromeric regions of the chromosomes across all composites but with variations in size and intensity of the FISH signals (Figures 1, 2 and 3). Comparative analysis of the composite karyotypes showed no homogeneity in the size of the CentC signals. The most striking variations occurred in chromosome pairs 5 and 9, with more pronounced polymorphisms in Composites 1, 2, 4 and 6 (Figures 3B, C, D and E). To further elucidate the number, size, location and distribution of 5S rDNA and the DNAsat probes across composite maize karyotypes analyzed here, idiograms were schematically constructed for each sample (Figure 3). The mapping showed that the K180 probe displayed the most striking variations in chromosome pairs 5, 6 and 9, due to the absence, presence and size of the FISH signals. Moreover, heteromorphisms were detected in chromosome pairs 1 (Composites 1 and 4), 2 (Composite 6), pair 4 of Composite 4, pair 5 (Composites 6 and 7), pair 6 of Composites 2, 6 and 7 and pair 10 of Composite 7 (Figure 3).

Composite 6 showed the highest heterogeneity, uniquely displaying K180 signals in both arms of chromosome 2 pair, heteromorphism (absence/presence) of K180 on pair 6, in addition to heteromorphisms for CentC in other composites (Figure 3E).

Comparative analysis of satDNA FISH probes (K180 and TR-1) versus maize flowering time revealed no clear trend or correlation between these parameters (Figure 4A). Our data showed no correlation between the increase or decrease in the number of FISH signals and the late or early flowering profile of the landraces of composite origin. Similarly, in the three NAM samples classified as intermediate (including B73), showed a smaller variation in K180 signals and a greater variation for TR-1, unlike the early samples where there was a large variation in K180 (especially in NAM) in relation to TR-1 sites. The late flowering samples showed a similar pattern to the intermediate samples, but with a marked variation for the in the Composite 7 (Figure 4A).

#### *Unclear relationship between 2C DNA content and maize flowering time*

Comparative DNA content analysis showed a variation from 5.18 pg (5,064.30 Mpb) in B73 to 5.86 pg (5,728.01 Mpb) in Composite 4 (Figure 4B and Supplementary Table 3). Statistical tests for normality (Shapiro-Wilk) indicated a Gaussian distribution, and the parametric analysis of variance (ANOVA) indicated that the values were significantly different ( $P = 0.0046$ ;  $F = 6.193$ ). In addition, the Tukey test also indicated that the DNA C values were significantly different. Composites 1 and 4 showed higher values compared to B73 (Figure 4B), but the other Composites (2, 4, 6 and 7) were statistically similar regarding DNA content. In Composites 1 and 4, with a greater number of K180 signals (17 signals), we observed the highest DNA 2C values (5.76 and 5.86 pg, respectively); however, these samples have different flowering times (early and late, respectively). Reference B73, with fewer K180 (12 FISH signals) and an intermediate flowering time had a 2C value of 5.18 pg. For TR-1, Composite 2

exhibited the highest number of FISH signals, a 2C value of 5.54 pg, and an early flowering time. Conversely, Composite 7, with the lowest number of TR-1 number, had a DNA 2C value of 5.52 pg and a late flowering time. Therefore, our results did not establish a reliable correlation between satDNA accumulation (K180 and TR-1), 2C DNA content, and flowering time.

*Mapping of the repetitive fraction in NAM maize pseudochromosomes*

Idiograms were constructed mapping and allocating the repetitive sequence positions into the chromosomes of NAM inbred lines, including two late-flowering, two intermediate-flowering, and two early-flowering lines, with B73 serving as an intermediate-flowering reference. Repeats with fewer monomer copies (<10 kbp) were predominantly located in the interstitial regions of most chromosome pairs, except for the short arms of chromosomes 3, 4, 7, 8, and 10 (Supplementary Figure 1). Larger satDNA arrangements (>10 kbp) showed a higher concentration in the chromosome pairs 5, 7, 8 and 9 (K180) and 6 (TR-1), mainly in the distal to subterminal regions (Supplementary Figure 1 and 2). Late-flowering genotypes exhibited a higher concentration of K180, with a tendency for K180 abundance to decrease in intermediate (including B73) and early-flowering genotypes (Supplementary Figure 2). In contrast, the distribution of TR-1 arrays varied independently of flowering time, particularly regarding the number of arrays with more monomers (larger blocks) (Supplementary Figure 2).

When comparing the genotypes concerning the distribution of the smaller blocks of both K180 and TR-1, these elements were predominantly located in the interstitial regions of all chromosome pairs, except in the short arms of chromosomes 3 and 10 (CML247), 10 (CML52), 7 and 10 (NC350), 4 and 10 (M37W), 3, 4, 7 and 10 (IL14H), 3, 7 and 10 (P39), although in this last genotype pair 7 did not show arrays in the long arm (Supplementary Figure 2). Regarding the relative number of repeats of these monomers in the NAM lineages, the B73 reference represented 0.94% of the genome, whereas in the CML52 genotype, TR-1 represented

0.17% of the genome. K180 loci were more concentrated on chromosomes 5 to 9 (Supplementary Figure 3A), whereas TR-1 arrays were heterogeneously distributed across chromosome pairs (Supplementary Figure 3B). The global number of DNAsat arrays in the pseudochromosomes was highest in the late flowering time accessions, decreased in intermediate flowering time genotypes, and was lowest in early flowering time genotypes.

*Distance between satDNA arrays and main genes related to maize flowering*

The estimated distances between the repetitive sequence arrays and the target genes, considering a minimum threshold of 175 kbp for each gene, revealed only arrays with a small number of monomers, ranging from 2 to 31 for K180 and from 2 to 19 for TR-1 (Supplementary Tables 5 and 6). For example, the M37W, CML52, NC350 and B73 lines had the smallest distances from TR-1 to the *zfl1* gene. The smallest annotated distances for K180 were 38,394 bp (IL14H lineage) and 39,461 bp (P39 lineage), both to the *dlfl* gene. In both cases, only two monomers were found. For the *zfl1* gene, the distance to the TR-1 arrays on chromosome 10 was relatively similar in all genomes. For the same gene, the distance to the K180 was similar only in the B73, CML247, P39 and IL14H genotypes with intermediate and late flowering. The distance of the closest K180 and TR-1 arrays to the 14 genes analyzed always showed a very low number of tandem repeats (<10 monomers). Given the small size and large distances from genes (always exceeding the 175 kbp limit), these arrangements are unlikely to explain the differences in flowering time based on the positional effects of heterochromatin.

*Density of Copia and Gypsy retrotransposons near flowering-related genes*

Regarding the proportion of Copia and Gypsy retrotransposons, the NC350 genotype contained more elements from the Copia lineages, while the M37W genotype had a higher number of Gypsy elements (Supplementary Figure 4). Overall, Copia insertions were more frequent than Gypsy insertions in these genes, except in B73, where Gypsy insertions

predominated. Retrotransposon density analysis, estimated at 50,000 bp upstream and downstream of each of the 14 genes involved in the flowering pathway, showed a large variation in the amount and diversity of fragments of non-autonomous LTR-RTs in adjacent regions. However, the biggest difference was the size of the genes and the number of insertions in each gene when we compared the late, intermediate, and early flowering lineages (Supplementary Figure 4 and 5). The B73 and CML247 genomes contained fewer LTR-RT fragments adjacent to the genes, while M37W accumulated more sequences of Copia and Gypsy, when compared to the other NAMs (Supplementary Figures 4 and 5). Two genes stood out for the relative number of RT-LTR insertions: *zfl2* and *mads1*. There were no RT-LTR insertions in the *zfl2* gene of the NC350 genotype, in contrast to the CML247 genotype, which had the highest number of insertions. The *mads1* gene showed no insertions in the CML247 and IL14H lineages, unlike the NC350, with the highest number of insertions. The gene with the lowest number of RT-LTR insertions in all genotypes was *zfl1*. In general, larger genes contained a greater number of LTR-RT insertions. Considering a minimum distance of kbp, we observe that, on average, 2.3% of the retrotransposons were inserted in these genic regions of maize.

## **Discussion**

*No clear relationship between satDNA accumulation, DNA C-values and flowering time in maize*

Comparative chromosome mapping using repetitive DNA probes revealed chromosomal variability in terms of knobs heterochromatin in composite maize landrace samples. This variation appeared to be more pronounced in comparison to the published data for other maize cultivars (Mondin et al. 2014; Realini et al. 2018; Silva et al. 2020). This is expected, as cytogenomic analyses focused on composite landraces derived from up to five parental lines. Comparison of all FISH mappings with the B73 reference showed no more or

less stable chromosome pairs regarding the presence, absence, size, and number of FISH signals. The observed variation is consistent with that reported by Kato et al. (2004) and Albert et al. (2018). Similarly, cytogenomic analyses of Argentine landraces (Realini et al. 2018) reported a high degree of intra- and interpopulation polymorphisms. Homozygotes and heterozygotes for K180 has also been reported in other cytogenomic studies, including accessions "Al Bandeirante" and "Jac-Duro" (Silva et al. 2020; Carvalho et al. 2021). *Zea diploperennis* also exhibited heterozygosity for K180 at positions K2S, K4L and K5L, although no heterozygotes were reported using the TR-1 probe (Xiong et al. 2005).

Attempts to correlate repetitive fraction heteromorphisms and polymorphisms with the amount of nuclear DNA have produced conflicting results. Realini et al. (2016) found a positive correlation between genome size and the percentage of heterochromatin in samples of Argentine landrace maize. However, in Guaraní landrace, DNA did not correlate with the number of K180 or the percentage of heterochromatin (Realini et al. 2021). Similarly, we did not observe a reliable correlation between genome size and the abundance of satDNA (K180 and TR-1) sequences in composite maize landraces. Our estimate of DNA content for the B73 reference showed 5.18 pg (5,064.3 Mbp), which is compatible with the 5.21 pg or 5,066.04 Mbp described in the literature (Tenailon et al. 2016). In comparison, the 2C values of the DNA of the composite maize we studied were slightly higher than those of B73 (from 5.4 to 5.86 pg). This variation was more limited than that reported for Argentine landrace maize, which ranged from 4.62 to 6.29 pg (Realini et al. 2016). Based on the comparison of genome sizes between tropical (5.39 to 6.24 pg) and temperate inbred maize, with a range of 5.12 to 5.83 pg, Jian et al. (2017) suggested that the genome size was also smaller in lines with shorter flowering time, although the number of K180 was higher. Our data showed that in Composite 6 samples with an early flowering time, the 2C value was 5.4 pg, but the number of K180 was the lowest. On the other hand, in Composite 7 that has the same

number of K180 as Composite 6, the genome size is smaller (5.52 pg), but the flowering time is late. These data indicate that the direct relationship between genome size and number of K180 does not apply to composite maize. Given that the DNA 2C values are within the range described for tropical varieties, we propose that direct relationships between repetitive fraction, nuclear DNA content and flowering time cannot be established without considering the fraction of transposable elements in each of the genome.

*Distance of the repetitive DNA arrays from genes involved in flowering*

The large diversity (number and size) in the satellite arrays observed after FISH assays and pseudochromosome annotation suggests dynamic rearrangements involving the accumulation of these satellite sequences across maize cultivars. On the other hand, large arrays tend to always occupy the distal chromosomal regions, with arrays >10 kbp concentrated on chromosomes 5, 7, 8 and 9 (K180) and 6 (TR-1), while arrays with fewer monomers (<10 kbp) appeared interstitially distributed in all pairs. This variation is also observed in the accumulation of satDNA sequences within and between genomes. The range from 0.94% in satellite sequences (B73) to 0.17% in CML52 is an example. Chen et al. (2023) reported that DNAsats (K180 and TR-1) were very abundant in the genome of Mo17 lineage, accounting for approximately 70% of the satellites.

Large arrays of satDNA monomers tend to organize heterochromatin sites, which may influence the genomes in different ways. Changes in heterochromatin content along chromosomes may affect the self-redistribution or accumulation/depletion of similar monomer arrays, as well as the changes in recombination, expression or silencing rate of surrounding genes, including suppress transposable elements via interference small RNAs (see Liu et al. 2020). This could be caused by unequal crossovers or illegitimate recombination, leading expansion or contraction of copy numbers. Despite the heteromorphisms and polymorphisms

associated with satDNA found here and reported in the literature, large satDNA arrays predominate in distal chromosomal regions, often arranging the K180. However, their precise role in gene regulation remains unclear. Therefore, future studies based on the analysis of the non-coding sequences could shed light on the effects of the satDNA fraction on the expression of genes located at long distances.

When genes are repositioned close to heterochromatin, they can be silenced by position effect variegation (PEV). Although gene activity can be maintained when the expression loop is moved away from the heterochromatic subnuclear domain, this does not always happen (see Jost et al. 2012). According to Jost et al. (2011), if the random distances are smaller than the actual distance of the gene to heterochromatin, it is not possible to ensure that heterochromatin sites act as gene silencers. In addition, the physical effect of position can occur when there is a greater concentration of heterochromatin close to the genes, within at least 175 kbp of genes (Voguel et al. 2009). Given the importance of flowering time-related genes and their proximity, our findings suggest the number of K180 and TR-1 monomers close these genes is very low to support the idea of a position effect.

*TEs adjacent to and intertwined with maize flowering route genes will be the subject of future studies*

The maize genome (>80%) is made up of transposable elements (TEs), mainly retrotransposons with long terminal repeats (Hufford et al. 2021; Chen et al. 2023). When TEs are located near or inserted into genes, they impact the gene performance, such as changes in transcriptional control and levels. Examples include the *tb1* gene expression in maize due to the insertion of TEs inhibiting tillering (Dong et al. 2019), as well as TEs insertions in the *ZmCCT9* and *ZmCCT10* genes favor earlier flowering (Yang et al. 2013; Huang et al. 2018). It has already been shown that an element like CACTA transposon inserted about 2 kb upstream

of the *ZmCCT* gene interfered with photoperiod sensitivity in maize (Yang et al. 2013).

Our estimates showed that most of the retrotransposon insertions were in the *zfl2* and *mads1* genes and that the gene with the fewest insertions was *zfl1*. Overexpression of the *mads1* gene results in an early flowering phenotype, highlighting its importance in flowering time (Alter et al. 2016). Our data show great variation in TE insertions in this gene, and that this could interfere with *zfl1* expression by inducing earliness, but further tests are needed to certify this effect. On the other hand, the *zfl2* gene acts as an upstream regulator of floral organogenesis, as well as in the phyllotaxis of the maize inflorescence (Bomblies et al. 2003). Similarly in *mads1*, TE insertions in *zfl2* were also highly variable. Insertion of transposons into the *zfl2* gene can result in defects in the architecture of the inflorescence and the transition from the vegetative to the reproductive phase (Bomblies et al. 2003). Although epigenetic effects resulting from the action of transposon elements are necessary, post-translational modifications of histones and the action of ncRNAs (non-coding RNA) from the TE fraction likely contribute to developmental regulation, responses to environmental stimuli and adaptation to the environment (Yu et al. 2020).

The maize genome has a large amount of TEs distributed along the chromosomes, including the gene-rich regions (Schnable et al. 2009). Our estimates and mappings showed that the interstitial regions contain 85% to 90% of TEs. The literature shows that gene-rich regions have 50% of these elements (Baucon et al. 2009). More than 86% of the genes have TEs inserted within 1 kb (Li et al. 2015), many of which are incomplete elements or fragments, as we also observed in the NAM strains. Considering the mapping of the repetitive sequences K180, TR-1 and TEs, about the 14 main genes listed as effectors in the metabolic pathways of flowering in maize, it was clear that the TEs insertions seem to be more relevant for future analyses associated with flowering than the DNAsat fraction.

## References

- Aguiar-Perecin ML, Fluminhan A, Santos-Serejo JAD, Gardingo JR, Bertão MR, Decico MJU, Mondin M. 2000. Heterochromatin of maize chromosomes: structure and genetic effects. *Genet Mol Biol* **23**: 1015-1019. doi:10.1590/S1415-47572000000400047
- Albert PS, Zhang T, Semrau K, Rouillard JM, Kao YH, Wang CJR, Danilova TV, Jiang J, Birchler JA. 2019. Whole-chromosome paints in maize reveal rearrangements, nuclear domains, and chromosomal relationships. *Proc Natl Acad Sci* **116**: 1679-1685. doi:10.1073/pnas.1813957116
- Alter P, Bircheneder S, Zhou LZ, Schlüter U, Gahrtz M, Sonnewald U, Dresselhaus T. 2016. Flowering Time-Regulated Genes in Maize Include the Transcription Factor *ZmMADS1*. *Plant Physiol* **172**: 389–404. doi:10.1104/pp.16.00285
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-10. doi:10.1016/S0022-2836(05)80360-2
- Ananiev EV, Phillips RL, Rines, HW. 1998a. A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons?. *Proc Natl Acad Sci* **95**: 10785-10790. doi:10.1073/pnas.95.18.10785
- Ananiev EV, Phillips RL, Rines, HW. 1998b. Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. *Proc Natl Acad Sci* **95**: 13073-13078. doi:10.1073/pnas.95.22.13073
- Best NB, Dilkes BP. 2022. Transcriptional responses to gibberellin in the maize tassel and control by DELLA domain proteins. *Plant J* **112**:493-517. doi:10.1111/tpj.15961
- Bilinski P, Albert PS, Berg JJ, Birchler JA, Grote MN, Lorant A, Quezada J, Swarts K, Yang J, Ross-Ibarra J. 2018. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genet* **14**: e1007162. doi:10.1371/journal.pgen.1007162
- Bombliès K, Wang RL, Ambrose BA, Schmidt RJ, Meeley RB, Doebley J. 2003. Duplicate FLORICAULA/LEAFY homologs *zfl1* and *zfl2* control inflorescence architecture and flower patterning in maize. *Development* **130**: 2385–2395. doi:10.1242/dev.00457
- Borem A, Miranda GV, Fritsche-Neto R. 2013. Melhoramento de plantas, 6<sup>a</sup> ed. Viçosa: Editora UFV.
- Braz GT, do Vale Martins L, Zhang T, Albert PS, Birchler JA, Jiang J. 2020. A universal chromosome identification system for maize and wild *Zea* species. *Chromosome Res* **28**: 183–194. doi:10.1007/s10577-020-09630-5
- Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, et al. 2009. The Genetic Architecture of Maize Flowering Time. *Science* **325**: 714-718. doi:10.1126/science.1174276
- Birchler JA; Yang H. 2021. The supernumerary B chromosome of maize: drive and genomic conflict. *Open Biol* **11**: 210197. doi:10.1098/rsob.210197
- Carvalho RF, Aguiar-Perecin MLR, Clarindo WR, Fritsche-Neto R, Mondin M (2022) A Heterochromatic knob Reducing the Flowering Time in Maize. *Front Genet* **12**: 799681. doi:10.3389/fgene.2021.799681
- Chen J, Wang Z, Tan K, Huang W, Shi J, Li T, Hu J, Wang K, Wang C, Xin B, et al. 2023. A complete telomere-to-telomere assembly of the maize genome. *Nat Genet* **55**:1221–1231.

doi:10.1038/s41588-023-01419-6

Dennis ES, Peacock WJ. 1984. Knob heterochromatin homology in maize and its relatives. *J Mol Evol* **20**: 341-350. doi:10.1007/BF02104740

Díez CM, Gaut BS, Meca E, Scheinvar E, Montes-Hernandez S, Eguiarte LE, Tenailon MI. 2013. Genome size variation in wild and cultivated maize along altitudinal gradients. *New Phytol* **199**: 264-276. doi:10.1111/nph.12247

Doležel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nat Protoc* **2**: 2233–2244. doi:10.1038/nprot.2007.310

Dong MY, Lei L, Fan XW, Li YZ. 2021. Analyses of open-access multi-omics data sets reveal genetic and expression characteristics of maize *ZmCCT* family genes. *AoB Plants* **13**: plab048. doi:10.1093/aobpla/plab048

Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15.

Dwivedi SL, Quiroz LF, Spillane C, Wu R, Mattoo AK, Ortiz R. 2024. Unlocking allelic variation in circadian clock genes to develop environmentally robust and productive crops. *Planta* **259**: 72. doi:10.1007/s00425-023-04324-8

Eickbush TH, Malik HS. 2007. Origins and evolution of retrotransposons. *Mob DNA II* 1111-1144. doi:10.1128/9781555817954.ch49

Feschotte C. 2023. Transposable elements: McClintock's legacy revisited. *Nat Rev Genet* **24**: 797–800. doi:10.1038/s41576-023-00652-3

Fourastié MF, Gottlieb AM, Poggio L, González GE. 2018. Are cytological parameters of maize landraces (*Zea mays* ssp. *mays*) adapted along an altitudinal cline?. *J Plant Res* **131**: 285–296. doi:10.1007/s10265-017-0996-3

Franklin KA, Davis SJ, Stoddart WM, Vierstra RD, Whitelam GC. 2003 Mutant analyses define multiple roles for phytochrome C in *Arabidopsis* photomorphogenesis. *Plant Cell* **15**: 1981–1989. doi:10.1105/tpc.015164

Ghaffari R, Cannon EK, Kanizay LB, Lawrence CJ, Dawe RK. 2013. Maize chromosomal knobs are located in gene-dense areas and suppress local recombination. *Chromosoma* **122**: 67-75. doi:10.1007/s00412-012-0391-8

Haberer G, Kamal N, Bauer E, Gundlach H, Fischer I, Seidel MA, Spannagl M, Marcon C, Ruban A, Urbany C et al. 2020. European maize genomes highlight intraspecies variation in repeat and gene content. *Nat Genet* **52**: 950–957. doi:10.1038/s41588-020-0671-9

Hassan AH, Mokhtar MM, El Allali A. 2024. Transposable elements: multifunctional players in the plant genome. *Front Plant Sci* **14**: 1330127. doi:10.3389/fpls.2023.1330127

Heuer S, Hansen S, Bantin J, Brettschneider R, Kranz E, Lörz H, Dresselhaus T. 2021. The Maize MADS Box Gene *ZmMADS3* Affects Node Number and Spikelet Development and Is Co-Expressed with *ZmMADS1* during Flower Development, in Egg Cells, and Early Embryogenesis. *Plant Physiol* **127**: 33–45. doi:10.1104/pp.127.1.33

Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, Liu J, Ricci WA, Guo T, Olson A, Qiu Y, et al. 2021. De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**: 655–662. doi:10.1126/science.abg5289

Hung H, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware

- D, Buckler ES, Doebley JF, et al. 2012. *ZmCCT* and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. *Proc Natl Acad Sci* **109**: E1913-E1921. doi:10.1073/pnas.1203189109
- Jafari F, Wang B, Wang H, Zou J. 2024. Breeding maize of ideal plant architecture for high-density planting tolerance through modulating shade avoidance response and beyond. *J Integr Plant Biol* **66**: 849-864. doi:10.1111/jipb.13603
- Jian Y, Xu C, Guo Z, Wang S, Xu Y, Zou C. 2017. Maize (*Zea mays* L.) genome size indicated by 180-bp K180 abundance is associated with flowering time. *Sci Rep* **7**: 5954. doi:10.1038/s41598-017-06153-8
- Jost KL, Bertulat B, Cardoso MC. 2012. Heterochromatin and gene positioning: inside, outside, any side?. *Chromosoma* **121**: 555–563. doi:10.1007/s00412-012-0389-2
- Liu P, Cuerda-Gill D, Shahid S, Slotkin RK. 2022. The epigenetic control of the transposable element life cycle in plant genomes and beyond. *Annu Rev Genet* **56**: 63-87. doi:10.1146/annurev-genet-072920-015534
- Kato A, Lamb JC, Birchler JA. 2004. Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci* **101**: 13554-13559. doi:10.1073/pnas.0403659101
- Kolawole AO, Menkir A, Blay E, Ofori K, Kling JG. 2018. Genetic advance in grain yield and other traits in two tropical maize composites developed via reciprocal recurrent selection. *Crop Sci* **58**: 2360-2369. doi:10.2135/cropsci2018.02.0099
- Kreiner JM, Wright SI. 2018. A less selfish view of genome size evolution in maize. *PLoS Genet* **14**: e1007249. doi:10.1371/journal.pgen.1007249
- Longley AE. 1941. Chromosome morphology in maize and its relatives. *Bot Rev* **7**: 263-289.
- McClintock B. 1929. Chromosome morphology in *Zea mays*. *Sci* **69**: 629-629, 1929.
- McClintock B, Kato TA, Blumenschein A. 1981. *Chromosome constitution of races of maize. its significance in the interpretation of relationships between races and varieties in the americas* (Vol. 31). Colegio de Pos graduados, Escuela Nacional de Agricultura.
- McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H, Sun Q, Flint-Garcia S, Thornsberry J, Acharya C, Bottoms C, et al. 2009. Genetic properties of the maize nested association mapping population. *Science* **325**: 737-740. doi:10.1126/science.1174320
- Meng X, Muszynski MG, Danilevskaya ON. 2011. The FT-like *ZCN8* gene functions as a floral activator and is involved in photoperiod sensitivity in maize. *Plant Cell* **23**:942-960. doi:10.1105/tpc.110.081406
- Mondin M, Santos-Serejo JA, Bertão MR, Laborda P, Pizzaia D, Aguiar-Perecin MLR. 2014. Karyotype variability in tropical maize sister inbred lines and hybrids compared with KYS standard line. *Front Plant Sci* **5**:544. doi:10.3389/fpls.2014.00544
- Monte E, Alonso JM, Ecker JR, Zhang Y, Li X, Young J, Austin-Phillips S, Quail PH. 2003. Isolation and Characterization of *phyC* Mutants in Arabidopsis Reveals Complex Crosstalk between Phytochrome Signaling Pathways. *Plant Cell* **15**: 1962–1980. doi:10.1105/tpc.012971
- Muszynski MG, Dam T, Li B, Shirbroun DM, Hou Z, Bruggemann E, Archibald R, Ananiev EV, Danilevskaya ON. 2006. *delayed flowering1* Encodes a Basic Leucine Zipper Protein That Mediates Floral Inductive Signals at the Shoot Apex in Maize. *Plant Physiol* **142**: 1523–

1536. doi:10.1104/pp.106.088815

Praça-Fontes MM, Carvalho CR, Clarindo WR, Cruz CD. 2011. Revisiting the DNA C-values of the genome size-standards used in plant flow cytometry to choose the “best primary standards”. *Plant Cell Rep* 30: 1183–1191. doi:10.1007/s00299-011-1026-x

Poggio L, Rosato M, Chiavarino AM, Naranjo CA. 1998. Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). *Ann Bot* 82: 107-115. doi:10.1006/anbo.1998.0757

Pulido M, Casacuberta JM. 2023. Transposable element evolution in plant genome ecosystems. *Curr Opin Plant Biol* 75: 102418. doi:10.1016/j.pbi.2023.102418

Realini MF, Poggio L, Cámara-Hernández J, González GE. 2016. Intra-specific variation in genome size in maize: cytological and phenotypic correlates. *AoB PLANTS* 8: plv138, 2016. doi:10.1093/aobpla/plv138

Realini MF, Poggio L, Cámara-Hernández J, González GE. 2018. Exploring karyotype diversity of Argentinian Guaraní maize landraces: relationship among South American maize. *PLoS One* 13: e0198398, 2018. doi:10.1371/journal.pone.0198398

Realini MF, Poggio L, Cámara-Hernández J, González GE. 2021. Genome size and repetitive sequences are driven by artificial selection on the length of the vegetative cycle in maize landraces from Northeastern Argentina. *Rdza* 72: e03542018. doi:10.1590/2175-7860202172004

Rhoades MM. 1942. Preferential segregation in maize. *Genetics* 27: 395. doi:10.1093/genetics/27.4.395

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *TiG* 16: 276-277. doi:10.1016/S0168-9525(00)02024-2

Roddy AB, Thérroux-Rancourt G, Abbo T, Benedetti JW, Brodersen CR, Castro M, Castro S, Gilbride AB, Jensen B, Jiang GF et al. 2020. The Scaling of Genome Size and Cell Size Limits Maximum Rates of Photosynthesis with Implications for Ecological Strategies. *Int J Plant Sci* 181: 75–87. doi:10.1101/619585v2

Romero Navarro JA, Willcox M, Burgueño J, Romay C, Swarts K, Trachsel S, Preciado E, Terron A, Delgado HV, Vidal V, et al. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat Genet* 49: 476-480. doi:10.1038/ng.3784

Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashv S, Bruggemann E et al. 2007. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci* 104: 11376-11381. doi:10.1073/pnas.0704145104

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Sci* 326: 1112-1115. doi:10.1126/science.1178534

Silva JC, Soares FAF, Sattler MC, Clarindo WR. et al. 2020. Repetitive sequences and structural chromosome alterations promote intraspecific variations in *Zea mays* L. karyotype. *Sci Rep* 10: 8866. doi:10.1038/s41598-020-65779-3

Shen X, Xiao B, Kaderbek T, Lin Z, Tan K, Wu Q, Yuan L, Lai J, Zhao H, Song W. 2023. Dynamic transcriptome landscape of developing maize ear. *Plant J* 116: 1856-1870.

doi:10.1111/tpj.16457

Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013–2015. <http://www.repeatmasker.org>.

Stitzer MC, Anderson SN, Springer SN, Ibarra JR. 2021. The genomic ecosystem of transposable elements in maize. *PLoS Genet* **17**: e1009768. doi:10.1371/journal.pgen.1009768

Souza TB, Parteka LM, Assis R, Vanzela ALL. 2022. Diversity of the repetitive DNA fraction in *Cestrum*, the genus with the largest genomes within Solanaceae. *Mol Biol Rep* **49**: 8785–8799. doi:10.1007/s11033-022-07728-z

Su H, Cao L, Ren Z, Sun W, Zhu B, Ma S, Sun C, Zhang D, Liu Z, Zeng H et al. 2024. *ZmELF6-ZmPRR37* module regulates maize flowering and salt response. *Plant Biotechnol J* **22**: 929–945. doi:10.1111/pbi.14236

Swentowsky KW, Gent JJ, Lowry EG, Schubert V, Ran X, Tseng KF, Harkess AE, Qiu W, Dawe RK. 2020. Distinct kinesin motors drive two types of maize neocentromeres. *Genes Dev* **34**: 1239–1251. doi:10.1101/gad.340679.120

Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, Wu S, Wang Y. 2023. SRplot: A free online platform for data visualization and graphing. *PLoS One* **18**: e0294236. doi:10.1371/journal.pone.0294236

Team, RC. 2020. R: A language and environment for statistical computing. *oalib* **8**. <https://www.r-project.org/>

Temsch EM, Koutecký P, Urfus T, Šmarda P, Doležel J. 2022. Reference standards for flow cytometric estimation of absolute nuclear DNA content in plants. *Cytometry Part A* **101**: 710–724. doi:10.1002/cyto.a.24495

Tenaillon MI, Manicacci D, Nicolas SD, Tardieu F, Welcker C. 2016. Testing the link between genome size and growth rate in maize. *PeerJ* **4**: e2408. doi:10.7717/peerj.2408

Vogel MJ, Pagie L, Talhout W, Nieuwland M, Kerkhoven RM, Van Steensel B. 2009. High-resolution mapping of heterochromatin redistribution in a *Drosophila* position-effect variegation model. *Epigenetics & chromatin* **2**: 1. doi:10.1186/1756-8935-2-1

Wang X, Wu L, Zhang S, Wu L, Ku L, Wei X, Xie L, Chenet Y. 2011. Robust expression and association of *ZmCCA1* with circadian rhythms in maize. *Plant Cell Rep* **30**:1261–1272. doi:10.1007/s00299-011-1036-8

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, PCapy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**: 973–982. doi:10.1038/nrg2165

Yang N, Wang Y, Liu X, Jin M, Vallebuena-Estrada M, Calfee E, Chen L, Dilkes BP, Gui S, Fan X et al. 2023. Two teosintes made modern maize. *Sci* **382**: eadg8940. doi:10.1126/science.adg8940

Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T et al. 2013. CACTA-like transposable element in *ZmCCT* attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. *Proc Natl Acad Sci* **110**: 16969–16974. doi:10.1073/pnas.1310949110

Yu J, Xu F, Wei Z, Zhang X, Chen T, Pu L. 2020. Epigenomic landscape and epigenetic regulation in maize. *Theor Appl Genet* **133**: 1467–1489. doi:10.1007/s00122-020-03549-5

Xiong ZY, Liu Y, He YG, Song YC, Li KX, He GY. 2005. Heterozygosity of Knob-Associated Tandem Repeats and Knob Instability in Mitotic Chromosomes of *Zea* (*Zea mays* L. and *Z. diploperennis* Iltis Doebley). *J Integr Plant Biol* **47**: 1345-1351. doi:10.1111/j.1744-7909.2005.00156.x

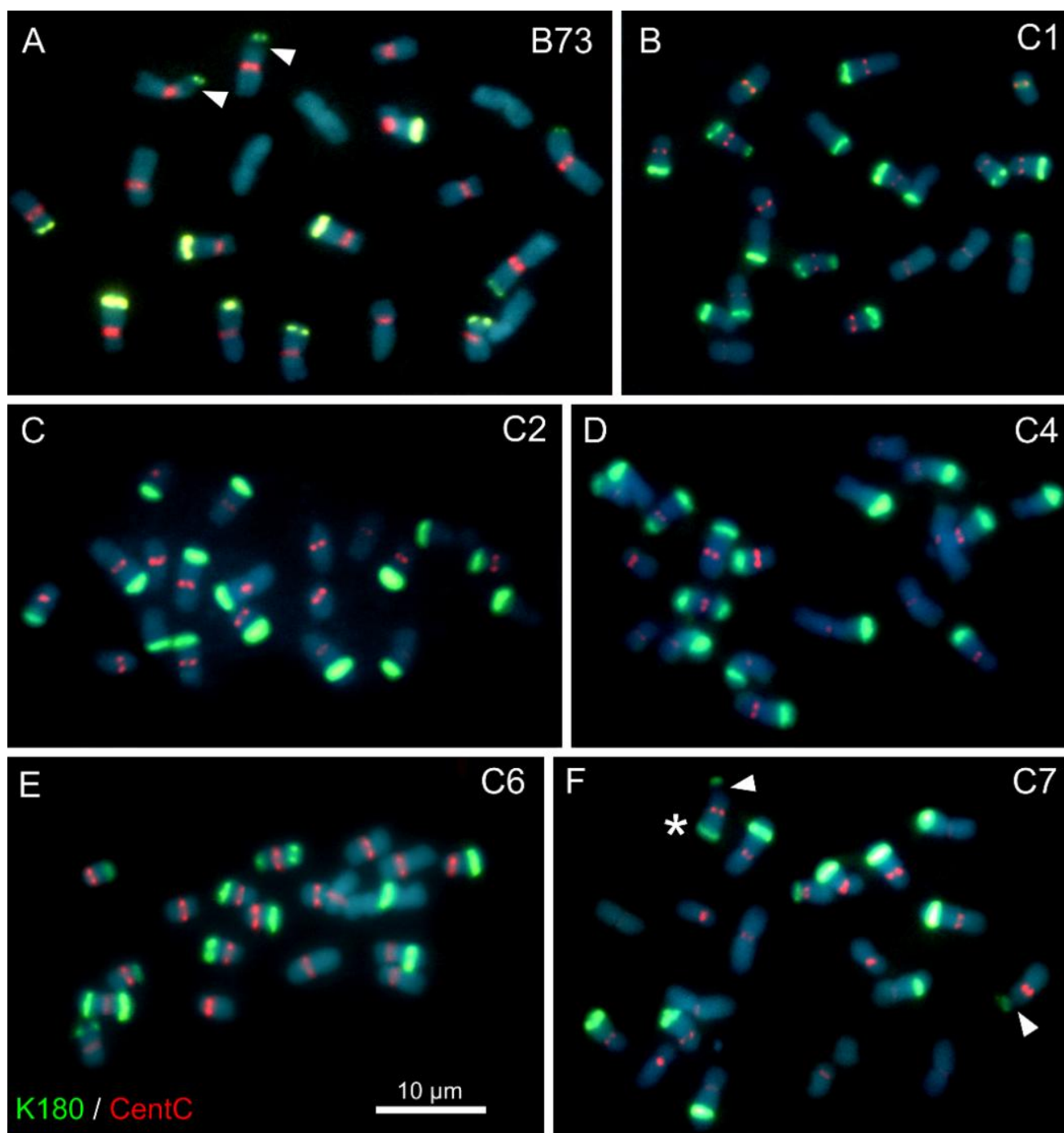
Zattera ML, Bruschi DP. 2022. Transposable elements as a source of novel repetitive DNA in the eukaryote genome. *Cells* **11**: 3373. doi:10.3390/cells11213373

## 6 CONCLUSÃO

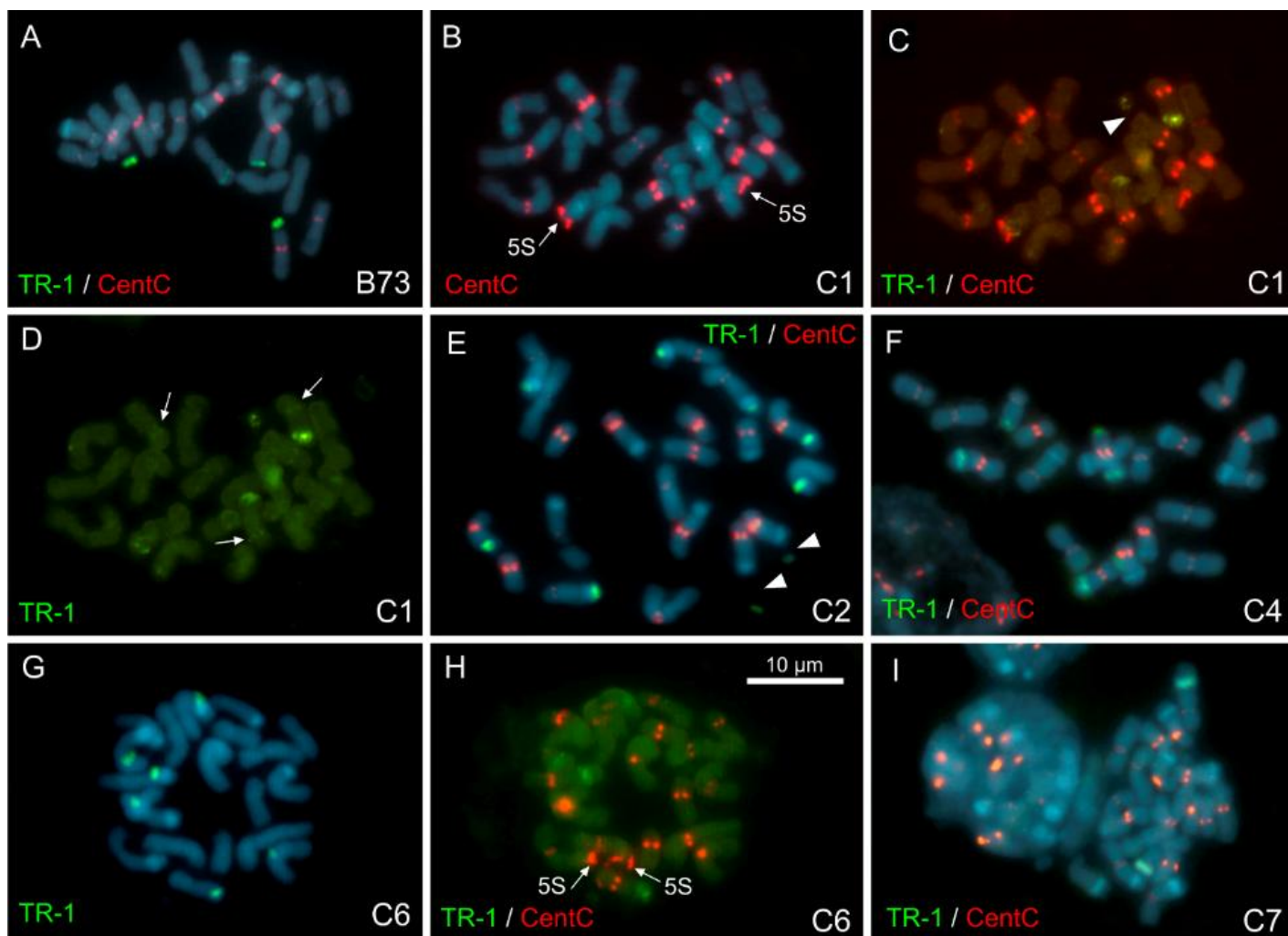
A quantidade, tamanho e localização dos DNAsats (K180 e TR-1) em associação com o tamanho do genoma não foi suficiente para fazermos uma correlação com o tempo de florescimento para as amostras de milhos compostos. Além disso, a anotação da coleção de sequências repetitivas próximas aos genes relacionados ao tempo de florescimento mostrou que não há uma sustentação quanto aos efeitos de posição de K180 e TR-1 sobre o tempo de florescimento em milho para as linhagens NAM. Além da distância, o baixo número de monômeros por arranjos das sequências de K180 e TR-1 ajudam a refutar essa hipótese. Contudo, a presença de fragmentos de retrotransposons acerca desses e inseridos nesses genes sugere que são necessários mais estudos de análises de transcriptomas desses genótipos com diferentes tempos de florescimento (precoce, intermediário e tardio). A transposição desses elementos nos genótipos de milho pode ser rastreadas, conjuntamente com os níveis de expressão de genes específicos de florescimento. Tais estudos podem clarificar a complexidade e efeitos da fração repetitiva de DNA e direcionar os programas de melhoramento focados no tempo de floração do milho.

## 7 FIGURAS

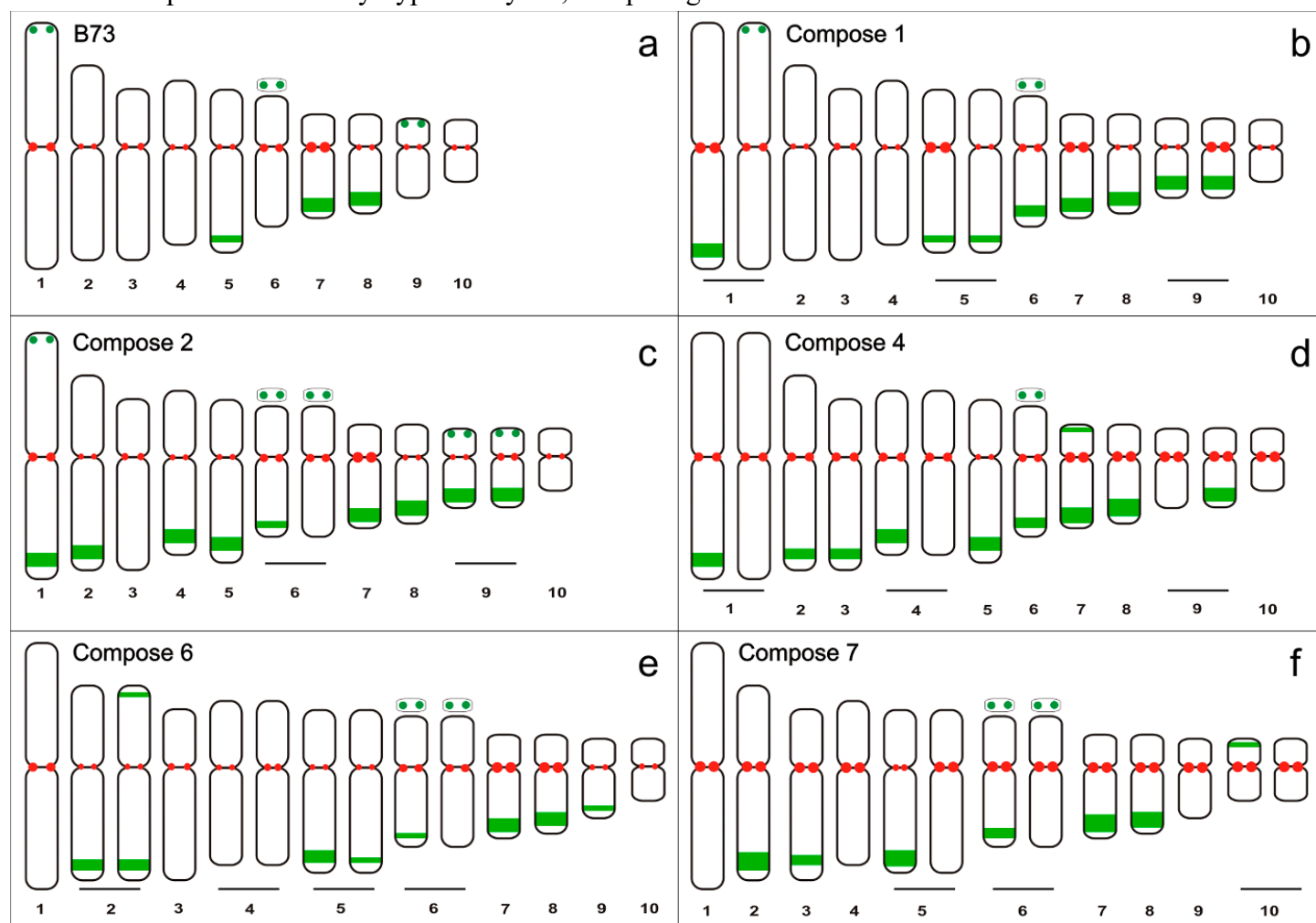
**Figure 1.** FISH on mitotic chromosomes of composite maize landraces using the K180 and CentC probes (green and red signals, respectively). The arrowheads in A and F indicate a secondary constriction. The asterisk indicates heteromorphism of chromosome pair 6. A) Mitotic metaphase of reference B73. B) Composite maize 1. C) Composite 2. D) Composite 4. E) Composite 6. F) Composite 7. Bar in E = 10  $\mu$ m.



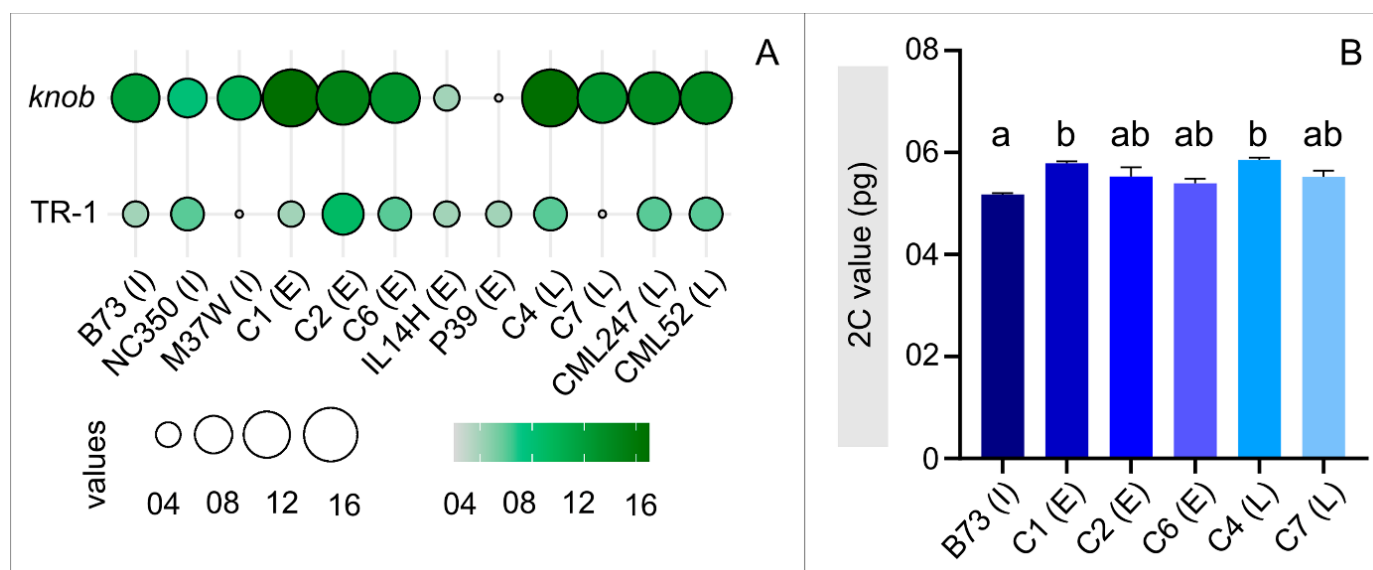
**Figure 2.** FISH using TR-1, 5S and CentC as probes in mitotic metaphase cells of different composite maize. A) Metaphase of reference B73 with TR-1 and CentC probes. B) Metaphase of Composite 1 with CentC and 5S rDNA probes. C) Metaphase of Composite 1 with TR-1 and CentC probes. D) Metaphase of Composite 1 with TR-1 probes and arrows highlighting the small arrangements. E) Metaphase of Composite 2 with TR-1 and CentC probes and arrowheads indicating the satellites of pair 6. F) Metaphase of Composite 4 with TR-1 and CentC probes. G) Metaphase of Composite 6 with the TR-1 probe. H) Metaphase of Composite 6 with TR-1, CentC and 5S DNAr probes. I) Metaphase of Composite 7 with TR-1 and CentC probes. Bar in E = 10  $\mu$ m.



**Figure 3.** Idiograms with the localization and distribution of DNAsat K180 and CentC sequences across the different composite maize karyotypes analyzed, comparing with the reference B73.



**Figure 4.** A) Number of K180 and TR-1 hybridization sites in composite and NAM maize. B) 2C DNA value (pg) of the composite maize. Different letters indicate significant differences between treatments ( $p < 0.05$ ). Mean  $\pm$  S.E.;  $n = 3$ .



## 8 SUPPLEMENTARY MATERIAL

**Supplementary Table 1.** Composite maize varieties and their respective reference. LMG: Laboratory for Genetic Improvement; C = Composites.

Samples	Origin	Parental / Genealogy	Grain color	Cycle	Flowering
C1	LMG	MC12 - Old common × Fine corncob	Yellow	Early	Late
		MC 19 - Mayan			
		MC21 - no name			
		MC23 - Purple straw			
C2	LMG	MC29 - Blushed	Yellow	Early	Late
		MC08 - Purple straw			
		MC10 - Cargill 408 × Agroceres			
		MC11 - Carioca			
C4	LMG	MC12 - Old common × Fine corncob	Yellow	Late	Late
		MC16 - Agouti tooth			
		MC36 - Yellow (Tião dos Borges)			
		MC37 - Yellow Grande			
C6	LMG	MC43 - Enchanted	White	Late	Intermediate
		MC50 - Gropires maize			
		MC55 - IAPAR 50			
		MC56 - Old maize			
C7	LMG	MC57 - Old Venglarek	White	Late	Late
		MC63 - White Buggy			
		MC66 - Vicente Huk white maize			
		MC67 - Eight Careers			
C7	LMG	MC69 - Aztec white	White	Late	Late
		MC70 - Aztec big white			
		MC72 - White of the Fence			
		MC73 - White of the North			
		MC79 - White with Purple Straw			

**Supplementary Table 2.** Characteristics of NAM strains and reference genomes. Accession codes at NCBI (<https://www.ncbi.nlm.nih.gov/>) and information from Buckler et al. (2009). Phenotypic characteristics of strains (<https://www.maizegdb.org/genome/assembly/Zm-CML52-REFERENCE-NAM-1.0>).

Lineage	Genome	DMB	DIS	TF	Phenotype
GCF_902167145.1	Zm-B73-NAM5.0	73.96	75.01	Intermediate	<i>"Stiff Stalk"</i>
GCA_902167175	M37W	79.51	82.32	Intermediate	Seasoned/Mix
GCA_902167005	NC350	82.09	82.35	Intermediate	Tropical
GCA_902166965	P39	65,15	66.82	Early	Sweet maize
GCA_902167065.1	IL14H	69.57	70.89	Early	Sweet maize
GCA_902166975	CML247	93.46	95.0	Late	Tropical
GCA_902373975	CML52	93.87	98.7	Late	Tropical

DMB = Days for the issue of the maize banner; DIS = Days for issuing stigmas;  
TF = Time for flowering

**Supplementary Table 3.** Fourteen genes related to flowering, their respective positions in the maize chromosomes and the accession code in MaizeGDB. L = long arm and S = short arm. CL = Chromosome location.

Genes	CL	Features	Access to MaizeGDB
<i>dwarf8</i>	1L	CO CO-like timing of CAB1 protein domain 14 - transcription factors for gibberellin signaling (Best and Dilke 2022)	Zm00001eb054480
<i>cct14</i>	2L	Regulation of photoperiodic flowering and salt stress response (Su et al. 2023)	Zm00001eb112080
<i>zfl2</i>	2S	<i>Zea floricaula</i> leafy2 - quantitative control of inflorescence phyllotaxy for maize (Bomblies et al. 2003)	Zm00001eb071990
<i>cca1</i>	4S	Circadian clock associated 1 - Acting on the maize circadian clock (Wang et al. 2011)	Zm00001eb172450
<i>phyC2</i>	5S	PhytochromeC2 - Inhibits flowering on short days by regulating to red light (Franklin et al. 2003; Monte et al. 2003)	Zm00001eb214800
<i>dlf1</i>	7L	delayed flowering1 - Acting in the floral transition (Muszynski et al. 2006)	Zm00001eb331880
<i>cct43</i>	7L	Co CO-Like Timing of CAB1 protein domain 43 - Acting on the maize circadian clock (Dong et al. 2021)	Zm00001eb331630
<i>rap2</i>	8L	rap2.7 orthologue - Negative regulator of the time of flowering (Salvi et al. 2007)	Zm00001eb355240
<i>pebp8</i>	8L	phosphatidylethanolamine-binding protein 8 - Coding for a florigen that integrates endogenous and environmental signals in maize (Jafari et al. 2023)	Zm00001eb353250
<i>mads1</i>	9L	Regulates the fate of the maize ear meristem (Shen et al. 2023)	Zm00001eb403750
<i>mads3</i>	9S	Co-expressed during the development of the ear and tassel in maize (Heuer et al. 2001)	Zm00001eb298680
<i>cct1</i>	10L	Co CO-Like Timing of CAB1 protein domain 1 - Regulator of later flowering (Hung et al. 2012)	Zm00001eb418700
<i>lhy1</i>	10L	Late hypocotyl elongation protein ortholog 1 - Acting on the maize circadian clock (Dwivedi et al. 2024)	Zm00001eb415760
<i>zfl1</i>	10L	<i>Zea floricaula</i> leafy 1 - regulator of the time of flowering (Bomblies et al. 2003)	Zm00001eb430240

**Supplementary Table 4.** Distribution of FISH signals in composite landrace maize lineages. DNA 2C values (Mpb), number and size (L=large and S=small) of FISH signals (FS) for K180, TR-1 and CentC probes. MFT = Maize-Flowering Time.

Samples	2C (Mpb)	K180 FS	TR-1 FS	CentC FS	MFT
B73	5,064.34	12 signals (4L e 8S)	4 signals (2L 2S)	16L 4S	Intermediate
C1	5,633.28	17 signals (10L e 7S)	4 signals (2L 2S)	9L 11S	Early
C2	5,418.12	15 signals (15L)	9 signals (4L 5S)	13L 7S	Early
C6	5,728.01	13 signals (9L e 4S)	6 signals (1L 5S)	8L 12S	Early
C4	5,278.34	17 signals (15L e 2S)	6 signals (6S)	18L 2S	Late
C7	5,401.00	13 signals (9L e 4S)	2 signals (2L)	12L 8S	Late

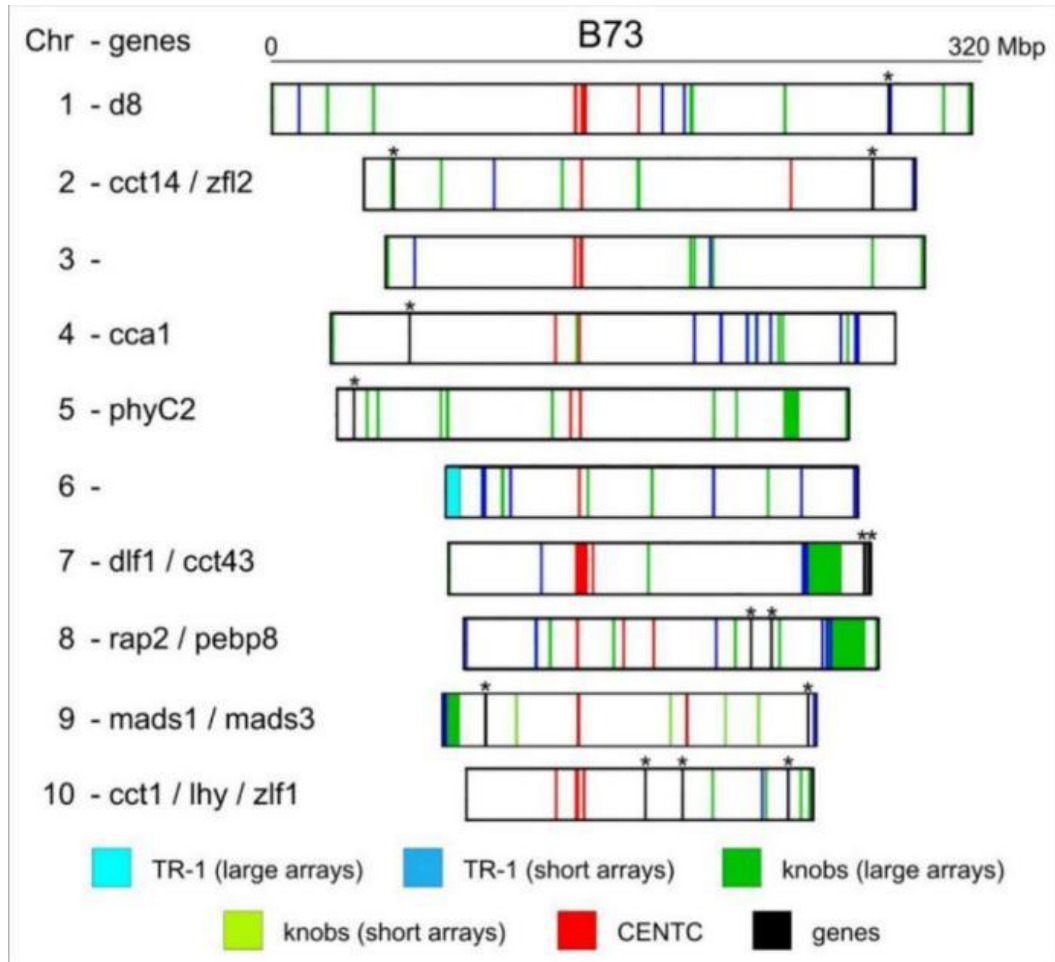
**Supplementary Table 5.** Distances less than 175 kbp from K180 to gene and number of K180 repeats at annotated position. *Zfl1* = *zea floriculata* / *leafy1*. *Dfl1*: *Delayed flowering 1*. CL = Chromosome location. L = long arm. NR = Number of Repeats.

Genome	CL	Gene	Distance (bp)	NR
B73	10L	<i>zfl1</i>	89.800	11
CML247	10L	<i>zfl1</i>	122.371	6
IL14H	10L	<i>zfl1</i>	125.005	31
P39	10L	<i>zfl1</i>	125.007	11
IL14H	7L	<i>dfl1</i>	38.394	2
P39	7L	<i>dfl1</i>	39.461	2

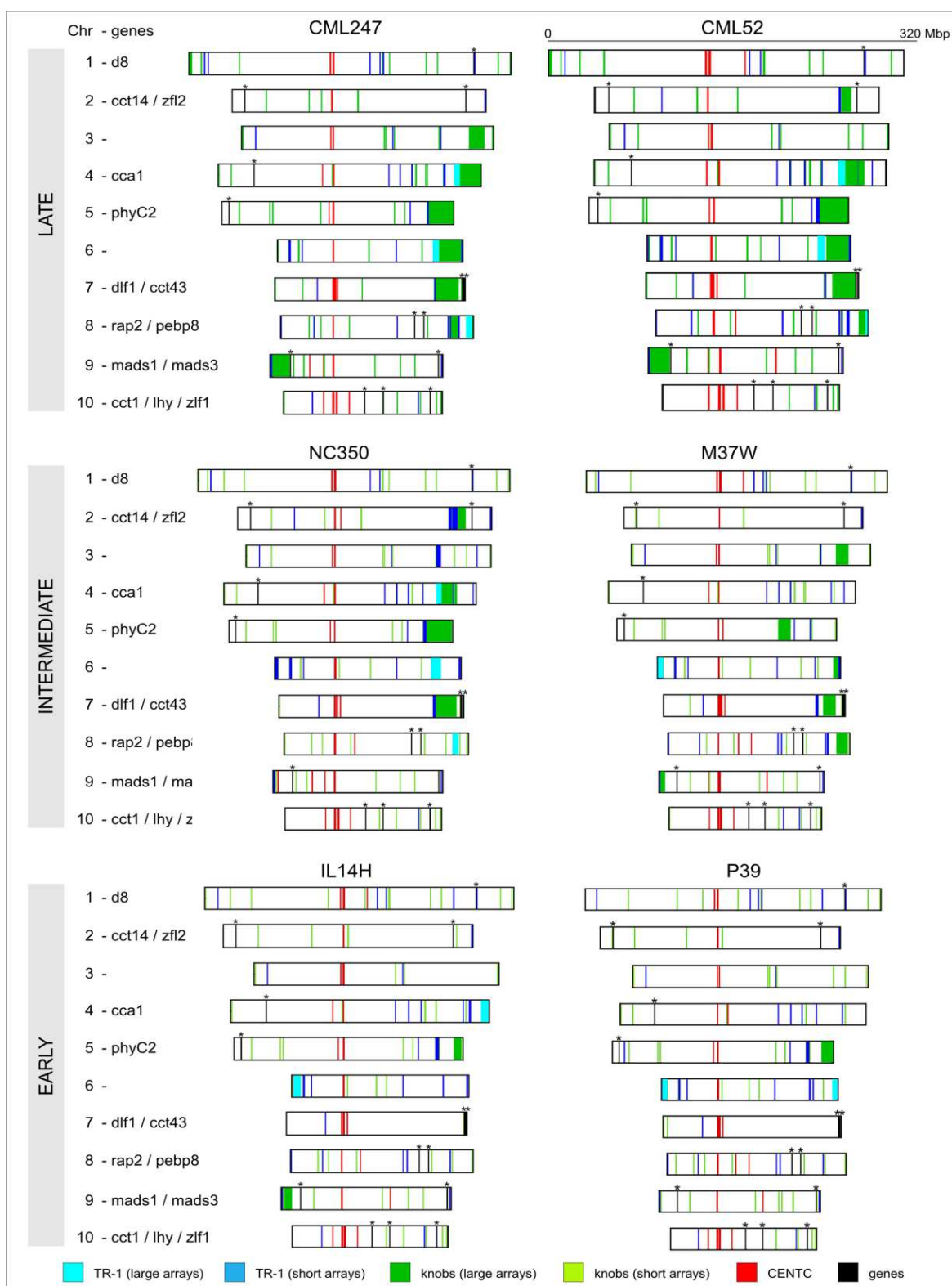
**Supplementary Table 6.** Distances less than 175 kbp from TR-1 to gene and number of TR-1 repeats at annotated position. *Zfl1* = *zea floriculata* / *leafy1*. CL = Chromosome location. L = long arm. NR = Number of Repeats.

Genoma	Chr	Gene	Distância (pb)	n° repetições
B73	10L	<i>zfl1</i>	90.941	4
NC350	10L	<i>zfl1</i>	79.130	2
CML52	10L	<i>zfl1</i>	120.241	1
CML247	10L	<i>zfl1</i>	124.238	1
M37W	10L	<i>zfl1</i>	124.975	1
IL14H	10L	<i>zfl1</i>	126.148	19
P39	10L	<i>zfl1</i>	126.150	19

**Supplementary Figure 1.** Idiograms of the pseudochromosomes of the NAM lines. The pair numbers are followed by the names of the genes related to the flowering time pathway they carry, and asterisks (\*) mark the position of each gene.

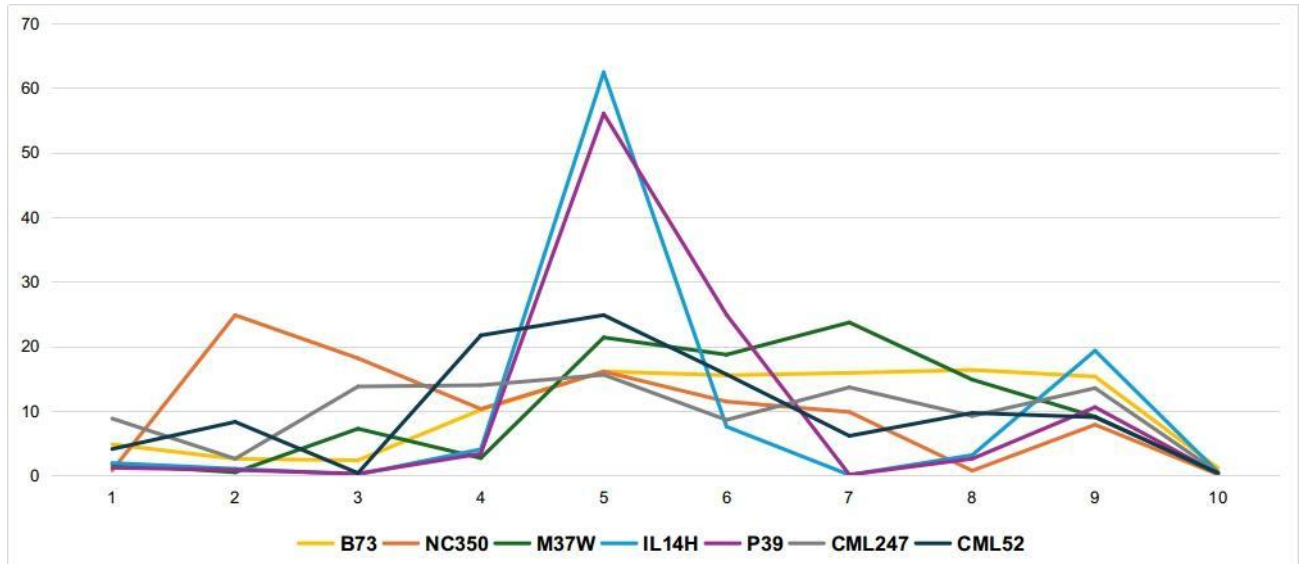


**Supplementary Figure 2.** Idiograms of the pseudochromosomes of the NAM lines (CML247, CML52, NC350, M37W, IL14H and P39). The pair numbers are followed by the names of the genes related to the flowering time pathway they carry, and asterisks (\*) mark the position of each gene. The pseudochromosomes have been centered according to the concentration of CentC arrays. Note that the late flowering lines accumulate more K180 than the others. However, in TR-1 the distribution of the blocks was more variable. Note also that the K180 on chromosome 9 (studied by Carvalho et al. 2022) is close to the *mads1* gene.

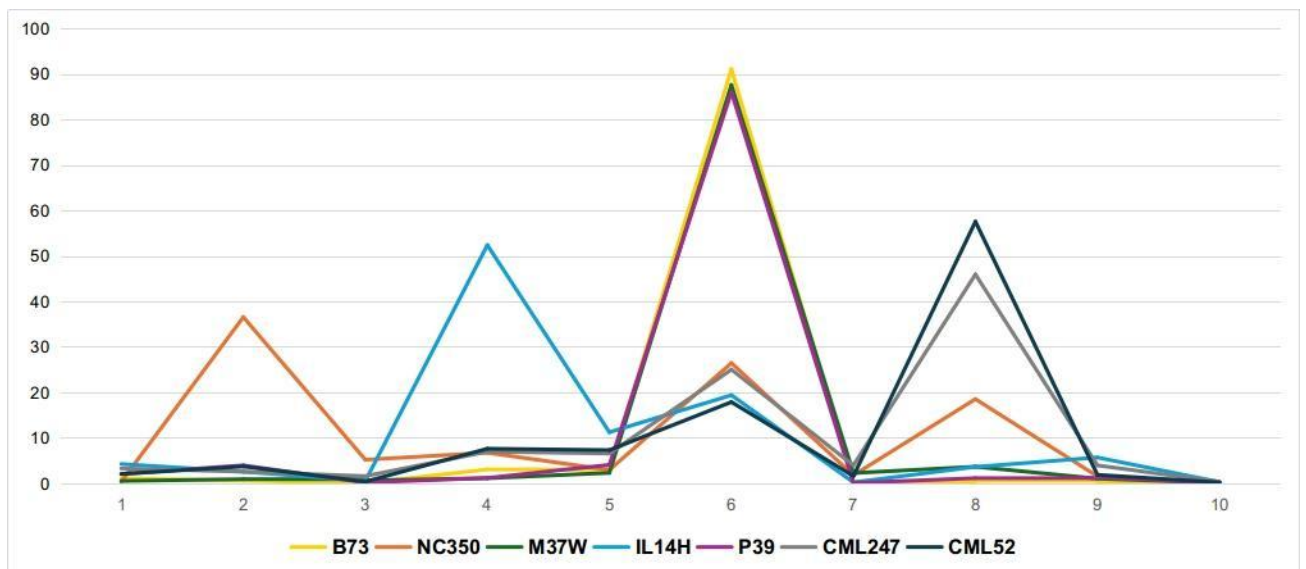


**Supplementary Figure 3.** Relative amount of K180 (A) and TR-1 (B) monomers in each pseudochromosome of NAM lineages.

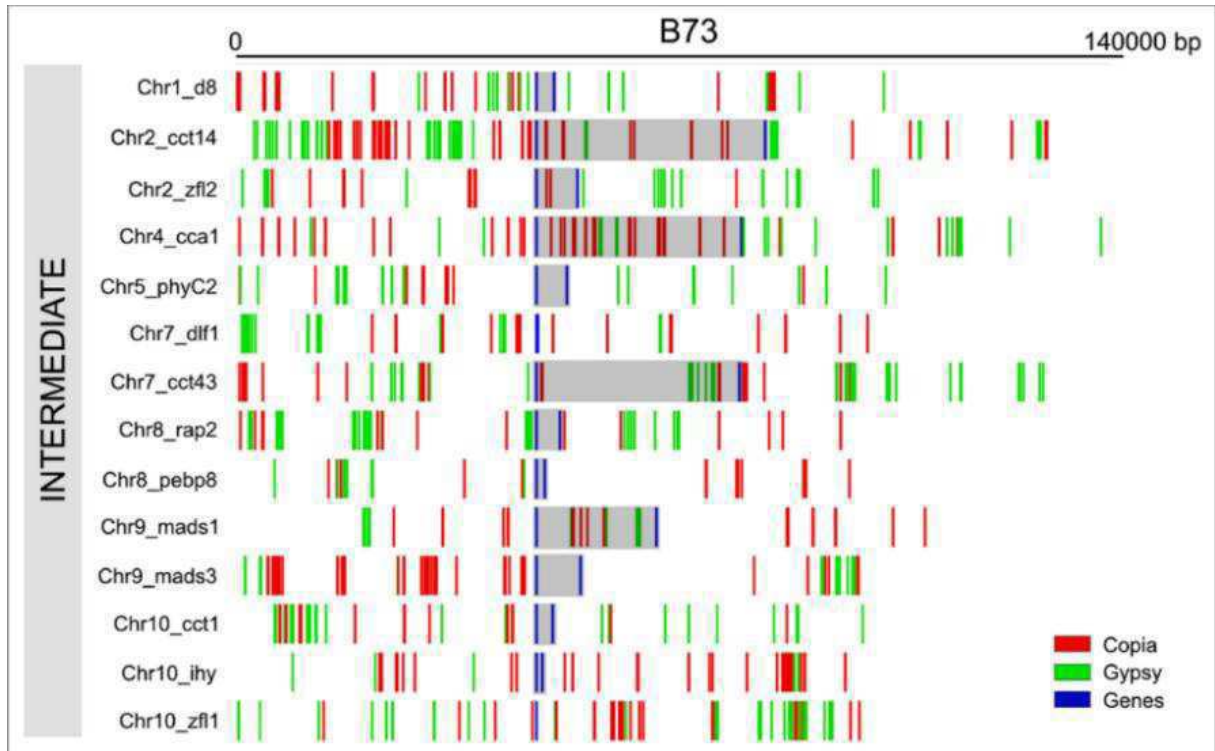
**A)**



**B)**



**Supplementary Figure 4.** Abundance of Copia and Gypsy retrotransposons in the flanking regions (50,000 bp upstream and downstream) of the 14 genes related to flowering time of NAM B73. The chromosome number is shown on the left, followed by the gene analyzed. The grey shaded region corresponds to the gene region annotated from the information contained in MaizeGDB. The two blue bars in the analyzed gene regions mark the start and end of each gene.



**Supplementary Figure 5.** Abundance of Copia and Gypsy retrotransposons in the flanking regions (50,000 bp upstream and downstream) of the 14 genes related to flowering time of NAM maize lines, namely late CML247 (A) and CML52 (B), intermediate NC350 (C) and M37W (D) and early IL14H (E) and P39 (F). The chromosome number is shown on the left, followed by the gene analyzed. The grey shaded region corresponds to the gene region annotated from the information contained in MaizeGDB. The two blue bars in the analyzed gene regions mark the start and end of each gene.

